December 2015

Heterogeneous Sensor Signal Processing for Inference with Nonlinear Dependence

Hao He
Syracuse University

ABSTRACT

Inferring events of interest by fusing data from multiple heterogeneous sources has been an interesting and important topic in recent years. Several issues related to inference using heterogeneous data with complex and nonlinear dependence are investigated in this dissertation. We apply copula theory to characterize the dependence among heterogeneous data.

In centralized detection, where sensor observations are available at the fusion center (FC), we study copula-based fusion. We design detection algorithms based on sample-wise copula selection and mixture of copulas model in different scenarios of the true dependence. The proposed approaches are theoretically justified and perform well when applied to fuse acoustic and seismic sensor data for personnel detection. Besides traditional sensors, the access to the massive amount of social media data provides a unique opportunity for extracting information about unfolding events. We further study how sensor networks and social media complement each other in facilitating the data-to-decision making process. We propose a copula-based joint characterization of multiple dependent time series from sensors and social media. As a proof-of-concept, this model is applied to the fusion of Google Trends (GT) data and stock/flu data for prediction, where the stock/flu data serves as a surrogate for sensor data.

In energy constrained networks, local observations are compressed before they are transmitted to the FC. In these cases, conditional dependence and heterogeneity complicate the system design particularly. We consider the classification of discrete random signals in Wireless Sensor Networks (WSNs), where, for communication efficiency, only local decisions are transmitted. We derive the necessary conditions for the optimal decision rules at the sensors and the FC by introducing a “hidden" random variable. An iterative algorithm is designed to search for the optimal decision rules. Its convergence and asymptotical optimality are also proved.

The performance of the proposed scheme is illustrated for the distributed Automatic Modulation Classification (AMC) problem. Censoring is another communication efficient strategy, in which sensors transmit only “informative" observations to the FC, and censor those deemed “uninformative". We design the detectors that take into account the spatial dependence among observations. Fusion rules for censored data are proposed with continuous and discrete local messages, respectively. Their computationally efficient counterparts based on the key idea of injecting controlled noise at the FC before fusion are also investigated.

In this thesis, with heterogeneous and dependent sensor observations, we consider not only inference in parallel frameworks but also the problem of collaborative inference where collaboration exists among local sensors. Each sensor forms coalition with other sensors and shares information within the coalition, to maximize its inference performance. The collaboration strategy is investigated under a communication constraint. To characterize the influence of inter-sensor dependence on inference performance and thus collaboration strategy, we quantify the gain and loss in forming a coalition by introducing the copula-based definitions of *diversity gain* and *redundancy loss* for both estimation and detection problems. A coalition formation game is proposed for the distributed inference problem, through which the information contained in the inter-sensor dependence is fully explored and utilized for improved inference performance.

HETEROGENEOUS SENSOR SIGNAL PROCESSING FOR INFERENCE WITH NONLINEAR DEPENDENCE

By

Hao He
B.S., Harbin Institute of Technology (HIT), China, 2010

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical and Computer Engineering

Syracuse University
December 2015

Acknowledgments

First and foremost, I would like to express my deepest gratitude and sincere thanks to my advisor, Prof. Pramod K. Varshney, for his guidance, support and encouragement during my doctoral study. I have been amazingly fortunate to have an advisor who gave me the freedom to explore on my own, and at the same time guidance to recover when my steps faltered. He has been influencing me with his enthusiasm towards research, as well as immense knowledge, and will keep guiding me in the future.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Biao Chen, Prof. Lixin Shen, Prof. Qi Cheng, Prof. Mustafa Gursoy and Prof. Fritz Schlereth for their insightful comments and suggestions.

I would like to thank all my fellow labmates in the Sensor Fusion Laboratory for their constant support in my research and life: Yujiao Zheng, Swarnendu Kar, Sid Nadendla, Raghed Bardan, Nianxia Cao, Aditya Vempaty, Sijia Liu, Bhavya Kailkhura, Shan Zhang, Qunwei Li, Prashant Khanduri, Swatantra Kafle and Pranay Sharma. Especially, discussions with Dr. Arun Subramanian and Dr. Satish Iyengar were very fruitful and triggered original ideas of part of this thesis.

Dr. Ruixin Niu guided me and encouraged me when I just started my program and some of his advice was beneficial throughout my entire doctoral study. Dr. Xiaojing Shen and Dr. Sora Choi have provided technical support for part of the work in this thesis.

My PhD research has been supported by US Army grants W911NF-09-1-0244, W911NF-14-1-0339 and W911-NF-13-2-0040 for which I am very grateful. The footstep data

used in Chapter 2 was provided by Army Research Laboratory (ARL). I would also like to thank Dr. Thyagaraju Damarla of ARL for providing very insightful suggestions on the research reported in this thesis

Most importantly, none of this would have been possible without the love and constant encouragement of my parents and my fiancee. I would like to thank them for always being there cheering me up and standing by me through the good times and bad.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

The problem of inferring events of interest by fusing data from multiple sensors has a wide variety of applications ranging from the surveillance of borders to disaster management. The inference tasks could consist of detecting an activity of interest or estimating some parameters, such as locations or tracks, that provide information for situational awareness. The sensors deployed in a given region of interest, in the most general setting, may consist of rather disparate and incommensurate modalities. In other words, with respect to the information content of the signals, sensors exhibit heterogeneity that can arise from a wide variety of causes. Another factor that influences the performance of such a multisensor system is the fact that the sensors observe different aspects of the same phenomenon, i.e., sensor observations are dependent. The nature of this dependence can be quite complex and nonlinear, especially in cases where the signal may propagate through a non-homogeneous medium. Inference in such a multisensor system is the major topic of this thesis.

Social media, facilitated by the growth of social networks, provides an easily accessible platform for users to share information and has resulted in the generation of unprecedented amounts of social media data that can be recorded and even monitored (such as, wall posts, clicks etc). This massive amount of social media data can be used by the signal process-

ing community, in combination with sensor data, for extracting information about unfolding events. This is expected to be beneficial in the military as well as civilian domains and provides us the motivation to investigate the convergence of sensor networks and social media in facilitating the data-to-decision making process and study how the two systems can complement each other for enhanced situational awareness. In networks with limited communication resources, local observations are usually compressed according to certain local rules, and only the compressed information is transmitted to the FC. In such distributed networks, the mechanism at sensors that is applied to reduce transmission can be quantization, censoring, etc. The existence of nonlinear inter-sensor dependence and heterogeneity of the network make the design of local rules and the fusion rule highly complex. We study the design of local and fusion rules in this thesis, which take inter-sensor dependence into consideration to improve inference performance. Local collaboration in the form of non-overlapping coalitions is also investigated to explore the dependence among sensors so that inference performance is improved to the largest extent under limited communication budget.

1.1 Background

Copula theory, which forms the basis of a lot of work in this thesis, is presented in this section.

1.1.1 Copula Theory

Copulas are parametric functions that couple univariate marginal distributions to a multivariate distribution. They can model the dependence among random variables with arbitrary marginal distributions. An important theorem that is central to the theory of copulas is Sklar's theorem (see Nelsen [75] for a detailed proof), which is stated below.

Theorem 1.1 (Sklar's Thoerem)**.** *Consider an N-dimensional distribution function F with marginal distribution functions $F_1, \ldots, F_N$. Then there exists a copula C, such that for all*

$x_1, \ldots, x_N$ *in* $[-\infty, \infty]$

$$F(x_1, x_2, \ldots, x_N) = C(F_1(x_1), F_2(x_2), \ldots, F_N(x_N)) \tag{1.1}$$

If F_n *is continuous for* $1 \leq n \leq N$*, then* C *is unique, otherwise it is determined uniquely on* $RanF_1 \times \ldots \times RanF_N$ *where* $RanF_n$ *is the range of cumulative distribution function (CDF)* F_n*. Conversely, given a copula* C *and univariate CDFs* $F_1, \ldots, F_N$*,* F *as defined in Eq.*(1.1) *is a valid multivariate CDF with marginals* $F_1, \ldots, F_N$*.*

Note that Eq. (1.1) implies that the copula function is a joint distribution of uniformly distributed random variables. As a direct consequence of Sklar's Theorem, for continuous distributions, the joint probability density function (PDF) is obtained by differentiating both sides of Eq. (1.1),

$$f(x_1, \ldots, x_N) = \Big(\prod_{n=1}^{N} f(x_n) \Big) c(F_1(x_1), \ldots, F_N(x_N)) \tag{1.2}$$

where, c is termed as the copula density and is given by,

$$c(\mathbf{u}) = \frac{\partial^N (C(u_1, \ldots, u_N))}{\partial u_1, \ldots, \partial u_N} \tag{1.3}$$

where, $u_n = F_n(x_n)$.

Thus, we can construct a joint density function with specified marginal densities by employing Eq. (1.2). Note that $C(\cdot)$ is a valid CDF and $c(\cdot)$ is a valid PDF for uniformly distributed random variables, u_n. The choice of a copula function to represent the joint statistics of the sensor observations is an important consideration here. Various families of copula functions exist in the literature [75]. However, which copula function should be used for a given case is not very clear as different copula functions may characterize different types of dependence behavior among the random variables [69]. A brief summary of some popularly used copula

functions is provided next.

1.1.2 Summary of Some Copula Functions

Copulas derived from distributions

Multivariate distribution functions specify dependence structures and copula functions can be derived from them. Two such copula functions are the Gaussian and the t copula functions that are derived from multivariate Gaussian and Student-t distributions respectively. Both specify dependence using the correlation matrix and are given as follows.

The Gaussian copula is defined as

$$C^G(\mathbf{u}|\Sigma) = \Phi_\Sigma(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_N)), \tag{1.4}$$

where, Φ_Σ denotes the multivariate normal CDF and Φ denotes the univariate normal CDF. Similarly, the t-copula is defined as

$$C^t(\mathbf{u}|\Sigma, \nu) = t_{\nu,\Sigma}(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_N)) \tag{1.5}$$

where, $t_{\nu,\Sigma}$ is the multivariate Student-t distribution with correlation matrix Σ and ν degrees of freedom ($\nu \geq 3$) and t_ν denotes the univariate Student-t distribution with ν degrees of freedom. It is common to set $\nu = 3$ to incorporate heavy tail dependence. As $\nu \to \infty$, the t copula approaches the Gaussian copula function. Both the Gaussian and the t copula functions belong to the elliptical family of copulas.

Table 1.1: Archimedean copula functions

Copula	Generator Function	Parametric Form
Clayton	$\frac{1}{\phi}\left(u^{-\phi}-1\right)$	$\left(\sum_{n=1}^{N} u_n^{-\phi}-1\right)^{-\frac{1}{\phi}},\ \ \phi\in[-1,\infty)\backslash\{0\}$
Frank	$-\log\frac{\exp\{-\phi u\}-1}{\exp\{-\phi\}-1}$	$-\frac{1}{\phi}\log\left(1+\frac{\prod_{n=1}^{N}[\exp\{-\phi u_n\}-1]}{\exp\{-\phi\}-1}\right), \phi\in\mathbb{R}\backslash\{0\}$
Gumbel	$-\ln u^{\phi}$	$\exp\left\{-\left(\sum_{n=1}^{N}(-\ln u_i)^{\phi}\right)^{\frac{1}{\phi}}\right\},\ \ \phi\in[1,\infty)$
Independent	$-\ln u$	$\prod_{n=1}^{N} u_n$

Archimedean copulas

Archimedean copulas, describing an m-variate CDF, are defined as follows,

$$C(\mathbf{u}|\phi)=\Psi^{-1}\left(\sum_{n=1}^{N}\Psi(u_n)\right) \tag{1.6}$$

where, $\Psi(\cdot)$ is referred to as the generator function and ϕ is the copula parameter specifying dependence. The Archimedean copulas considered in the chapters to follow are shown in Table 1.1 (see [75]).

1.1.3 Copulas and Measures of Dependence

An attractive feature of copulas is that nonparametric rank-based measures of dependence, such as Kendall's τ, can be expressed as expectations over the copula distribution. For independent pairs of random variables (X_1, Y_1) and (X_2, Y_2) having the same distribution as (X, Y), concordance is defined as the condition that $(X_1-X_2)(Y_1-Y_2)\geq 0$ and discordance is defined as the condition that $(X_1-X_2)(Y_1-Y_2)<0$. Kendall's τ is defined to be the difference between the probabilities of concordance and discordance:

$$\tau \triangleq P[(X_1-X_2)(Y_1-Y_2)\geq 0]-P[(X_1-X_2)(Y_1-Y_2)<0].$$

Nelsen has proved the relationship in (1.7) for a copula, C, and random variables $X \sim f_X(x), Y \sim f_Y(y)$ [75, p. 161], i.e.,

$$\tau(\boldsymbol{\phi}) = 4\,\mathbb{E}[C_\phi(F_X(x), F_Y(y))] - 1. \tag{1.7}$$

This relationship allows τ to be expressed in terms of the dependence parameter of the copula, C (Σ for the elliptical copulas and ϕ for the Archimedean copulas in Table 1.1). For the case of elliptical copulas, parametrized by the matrix $\Sigma = [\rho_{mn}]$,

$$\rho_{mn} = \sin\left(\frac{\pi\tau_{mn}}{2}\right), \tag{1.8}$$

where τ_{mn} is the Kendall's τ evaluated for the pair $(U_m, U_n) = (F_{X_m}(\cdot), F_{X_n}(\cdot))$. The sample estimate of Kendall's τ, for N observations, can be calculated as the ratio of the difference in the number of concordant pairs, c, and discordant pairs, d, to the total number of pairs of observations, i.e.,

$$\hat{\tau} = \frac{c-d}{c+d} = \frac{c-d}{\binom{N}{2}} \tag{1.9}$$

Typically, the value of the dependence parameter is not known *a priori*, and $\boldsymbol{\phi}$ needs to be estimated, e.g., using maximum likelihood estimation (MLE). On the other hand, (1.9) and (1.7) imply that Kendall's τ can be used for calculating computationally efficient estimates of $\boldsymbol{\phi}$.

1.2 Literature Review

Multisensor signal processing may be viewed as a subset of the broader field of information fusion. *Centralized* formulations, where raw observations are available at the processing unit or *fusion center*, for several inference tasks are well known and available in standard textbooks [14, 61, 107]. *Distributed* inference, on the other hand, relies on the availability of a

network that can either transmit local inferences/quantized measurements to the fusion center or arrive at a consensus solution by locally exchanging compressed/quantized information. While research in this area has forked in various directions, the problems addressed can be categorized as either distributed detection [114] or decentralized estimation (e.g., see [76, 78, 84] and references cited therein).

This section reviews recent progress that has been made in the field of multisensor signal processing, and focuses on developments where dependence information plays a significant role in the design. The aim of the discussion, as presented, is to motivate the relevance of our current research. The emphasis on dependence notwithstanding, the literature is quite extensive, and instead of being exhaustive, we concentrate on highlighting newer developments.

1.2.1 Dependence as Covariance

Modeling dependence as a covariance matrix (or equivalently a correlation matrix) is arguably one of the most popular ways of characterizing dependence. It defines the dependence of jointly normal random variables and describes the linear dependence between random variables that possess a finite second moment. In the centralized paradigm, it is used extensively to model the dependency information for array signal processing applications, especially where it is reasonable to assume linearity of the medium of signal propagation. The most recent technologies where these concepts of array signal processing have been applied are MIMO radar [67] and joint blind source separation (JBSS) [5]. In MIMO radar, several antenna elements are used to transmit multiple probing signals that may be correlated or uncorrelated with one another. While traditional blind source separation problems are formulated using a single dataset, JBSS formulations are useful when analyzing multiple datasets as a group. An example of this is the separation of speech and audio signals in multiple frequency bands. The fusion of EEG with fMRI data for the detection of schizophrenia is discussed by Correa et al. [26] where the brain tissue is modeled as a mixing channel, and hence the information fusion problem is posed as

a JBSS problem and is solved using an approach based on multivariate canonical correlation analysis [62]. Canonical correlation analysis (CCA) has also been used for audio-video fusion: Slaney and Covell [93] use CCA to measure the synchrony between acoustic features and video frames, while Kidron et al. [63] consider a CCA based approach to determine pixels in images that exhibit maximal correlation with the acquired audio signal.

Optimal schemes for distributed inference with correlated observations has also been a topic of considerable interest. In the case of distributed detection, it has been shown that the likelihood ratio based quantizer, which was optimal under the assumption of conditional independence, is no longer optimal when correlation is taken into account. Examples of the consequent loss in performance are provided by Aalo and Viswanathan [1]. In fact, earlier work by Tsitsiklis and Athans [111] has shown that the distributed detection problem with dependent observations is NP-complete. One way to get past the computational intractability is to assume some prior knowledge about the joint statistics: Drakopolous and Lee [33] examine the fusion rule for distributed detection under dependence by considering that the correlation coefficient is known, whereas Kam et al. [58] use the Bahadur-Lazarsfeld expansion of probability density functions. Willett et al. [123] study the problem of distributed detection of a mean shift in correlated Gaussian noise and establish how the nature of correlation affects the optimum fusion rule. They conclude that even for a simple two-sensor and linear correlation formulation the distributed detection problem "exhibits apparently very complicated behavior."

The decentralized estimation problem with correlated observations has been studied by Fang and Li [39]. They consider a power constrained wireless sensor network [104] and examine power allocation for spatially correlated sensor observations. Each sensor transmits a possibly nonlinear function of the parameter of interest, θ, that is corrupted by additive, correlated Gaussian noise. Bandwidth constrained formulations requiring quantized transmissions to the fusion center are also considered by Ribeiro and Giannakis [84]. However, they consider a linear observation model, with θ being deterministic but unknown, and hence the sensor

observations are conditionally independent. Krasnopeev et al. [64] present a distributed estimation scheme for the problem $x_n = \theta + w_n$, where x_n is the measurement of sensor n and the noise w_n is a multivariate Gaussian correlated spatially across sensors. The covariance is assumed to be known at the fusion center. We note that all these problems are considered to be distributed since each local sensor transmits some local estimate of θ, which in its simplest form is the noise corrupted parameter itself. These formulations do not consider local, internode communication; the implications of this local communication aspect have been recently investigated by Kar et al. [60].

1.2.2 Nonlinear Dependence: Nonparametric Approach

Nonparametric approaches to multisensor signal processing have been very popular in applications where it is infeasible to model *a priori* the complex dependencies that may exist between the signals/features acquired by the sensors. These methods, in essence, estimate or learn the joint distribution across sensor measurements directly from the data.

Machine learning techniques fall under this framework and are applicable largely when it is feasible to control environment variables in such a way that a representative training dataset may be collected. While this is apparently a stringent requirement, often with some preprocessing, a significant amount of information can be extracted from sensor observations. This has led to the successful application of machine-learning techniques for a wide variety of problems. Learning based methodologies have been successfully applied to multibiometric systems [12, 85]. Multibiometric systems achieve superior personnel identification performance by fusing information from two or more biometric modalities. The learning-based approach has also been popular for solving several object classification tasks [57, 73] and have traditionally focused on security and surveillance applications [66, 132]. Recently, challenges unique to emerging technologies such as ubiquitous and human-centered computing have led to new research in areas such as object tracking and affect recognition [125, 128].

When viewed from an information fusion perspective, nonparametric designs offer tangible advantages over methods described in Section 1.2.1. Fusion of heterogeneous or multimodal information is possible since disparate modalities are not constrained to a multivariate normal approximation. For example, Butz and Thiran [18] use the mutual information and joint entropy between audio and video data as a measure of dependence; the joint density required for the computation of these quantities is estimated from the data using the nonparametric Parzen's estimator [119]. Graphical models such as Bayesian networks generalize hidden Markov models and have also been successfully used for audio-visual tracking [11, 28, 57]. Algorithms for distributed fusion using graphical models have been described by Çetin et al. [19].

1.2.3 Nonlinear Dependence: Copula-based Approach

Recall from Section 1.1.1 that copulas are parametric functions that couple univariate marginal distribution functions to the corresponding multivariate distribution function. A copula-based formulation is attractive because the spatial correlation among sensor observations can get manifested in several different, potentially non-linear ways and many families of copula functions have been specified in the literature to address this issue. Further, while nonparametric formulations are known to converge to the true distribution asymptotically, they also suffer from scalability issues stemming from the curse of dimensionality. Recently, considerable progress has been made in the study of copulas and their applications in statistics. The usage of copulas is widespread in the fields of econometrics and finance [25] and they are beginning to be used in the signal and image processing context [29, 52, 71, 103].

Iyengar et al. [51] have investigated the general framework of copula-based detection of a phenomenon being observed jointly by heterogeneous sensors. They quantify the performance loss due to copula misspecification and demonstrate that a detector using a copula selection scheme based on area under the receiver operating characteristic (ROC) can provide significant improvement over models assuming independence. Their results on a NIST multibiomet-

ric dataset show that the copula based approach is versatile and can fuse not only heterogeneous sensor measurements, but can also be applied to fuse different algorithms. Sundaresan et al. [100] consider the case of distributed detection and derive the optimum fusion rules for a Neyman-Pearson detector. Sundaresan and Varshney [99] also design and analyze the performance of a copula-based estimation scheme for the localization of a radiation source.

1.3 Main Contributions and Organization

The main contributions of the research results presented in this dissertation to the signal processing and information fusion literature, are as follows:

In Chapter 2, we investigate the detection problem using heterogeneous sensor data, where observations from disparate sensors may be conditionally dependent. We use copula theory to construct a valid joint distribution, to describe the dependence among these heterogeneous data. We consider the dependence to be “complex", where the “complexity" is assumed to include two cases, one is when the dependence structure is non-stationary, the other is when the true dependence is beyond the description of a single copula. For the scenario of non-stationary dependence, we propose a sample-wise copula selection rule and theoretically justify its utility. For the other scenario, we use a mixture of copulas to approach the true underlying dependence and use expectation maximization (EM) algorithm to solve for the unknown parameters in the model. Both theoretical proofs and simulation results are provided to show that the mixture of copula model results in a better detection performance compared with previously used single copula method. We also apply our proposed approaches to personnel detection using real footstep data collected by acoustic and seismic sensors.

The access to the massive amount of social media data provides a unique opportunity to the signal processing community for extracting information that can be used to infer about

unfolding events. It is desirable to investigate the convergence of sensor networks and social media in facilitating the data-to-decision making process and study how the two systems can complement each other for enhanced situational awareness. In Chapter 3, we propose a copula-based joint characterization of multiple dependent time series from sensors and social media. As a proof-of-concept, this model is applied to the fusion of Google Trends (GT) data and stock price data of Apple Inc. for prediction, where the stock data serves as a surrogate for sensor data. We also apply our model to the fusion of GT data and Influenza-Like Illness (ILI) data from Centers for Disease Control and Prevention (CDC) for flu prediction. Superior prediction performance of our method is demonstrated for both problems, by taking the non-linear dependence among social media data and sensor data into consideration.

In Chapter 4, we consider the distributed classification of discrete random signals in wireless sensor networks (WSNs). Observing the same random signal makes sensors' observations conditionally dependent which complicates the design of distributed classification systems. In the literature, this dependence has been ignored for simplicity although this may significantly affect the performance of the classification system. We derive the necessary conditions for the optimal decision rules at the sensors and the fusion center (FC) by introducing a "hidden" random variable. Furthermore, we introduce an iterative algorithm to search for the optimal decision rules, the convergence of which is also proved. The proposed scheme is applied to a distributed Automatic Modulation Classification (AMC) problem. It is shown to attain superior performance in comparison with other approaches which disregard the inter-sensor dependence.

In Chapter 5, we consider a distributed detection problem for a censoring sensor network where each sensor's communication rate is significantly reduced by transmitting only "informative" observations to the Fusion Center (FC), and censoring those deemed "uninformative". Our focus is on designing the fusion rule under the Neyman-Pearson (NP) framework that takes into account the spatial dependence among observations. Two transmission scenarios are

considered, one where uncensored observations are transmitted directly to the FC and second where they are first quantized and then transmitted to further improve transmission efficiency. Copula-based Generalized Likelihood Ratio Test (GLRT) for censored data is proposed with both continuous and discrete messages received at the FC corresponding to different transmission strategies. We address the computational issues of the copula-based GLRTs involving multidimensional integrals by presenting more efficient fusion rules, based on the key idea of injecting controlled noise at the FC before fusion. Although, the signal-to-noise ratio (SNR) is reduced by introducing controlled noise at the receiver, simulation results demonstrate that the resulting noise-aided fusion approach based on adding artificial noise performs very closely to the exact copula-based GLRTs. Copula-based GLRTs and their noise-aided counterparts by exploiting the spatial dependence greatly improve detection performance compared with the fusion rule under independence assumption.

We consider the problem of collaborative inference in a sensor network with heterogeneous and statistically dependent sensor observations. Each sensor aims to maximize its inference performance by forming a coalition with other sensors and sharing information within the coalition. In Chapter 6, the formation of non-overlapping coalitions with statistically dependent sensors is investigated under a communication constraint. We apply a game theoretical approach to fully explore and utilize the information contained in the spatial dependence among sensors to maximize individual sensor performance. Before formulating the distributed inference problem as a coalition formation game, we quantify the gain and loss in forming a coalition by introducing the concepts of *diversity gain* and *redundancy loss* for both estimation and detection problems. These definitions, enabled by the statistical theory of copulas, allow us to characterize the influence of statistical dependence among sensor observations on inference performance and collaboration strategy. An iterative algorithm based on merge-and-split operations is proposed for the solution and the stability of the proposed algorithm is analyzed. Numerical results are also provided for illustration.

Finally, in Chapter 7, we summarize the findings and results of this dissertation. Several directions and ideas for future research are also presented.

1.4 Bibliographic Note

Part of the work presented in this dissertation has appeared in the following publications:

1. Hao He, Arun Subramanian, Pramod K. Varshney, and Thyagaraju Damarla, "Fusing Heterogeneous Data for Detection Under Non-stationary Dependence," in *Proc. 15th International Conference on Information Fusion*, 2012.

2. Hao He, Arun Subramanian, Xiaojing Shen, and Pramod K. Varshney, "A Coalitional Game for Distributed Estimation in Wireless Sensor Networks," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013.

3. Hao He, and Pramod K. Varshney, "Distributed Detection with Censoring Sensors under Dependent Observations," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014.

4. Hao He, Sora Choi, Pramod K. Varshney, and Wei Su, "Distributed Classification under Statistical Dependence with Application to Automatic Modulation Classification," in *Proc. 18th International Conference on Information Fusion*, 2015.

5. Hao He, Arun Subramanian, Sora Choi, Pramod K. Varshney, and Thyagaraju Damarla, "Social Media Data Assisted Inference with Application to Stock Prediction," to appear in *Proc. IEEE Asilomar Conference on Signals Systems and Computer*, 2015.

6. Hao He, and Pramod K. Varshney, "Fusing Censored Dependent Data for Distributed Detection," *Signal Processing, IEEE Transactions on*, vol 63, no. 16, pp. 4385-4395, Aug. 15, 2015.

7. Hao He, and Pramod K. Varshney, “A Coalitional Game for Distributed Inference in Sensor Networks with Dependent Observations," to appear in *Signal Processing, IEEE Transactions on*.

CHAPTER 2
CENTRALIZED DETECTION UNDER COMPLEX DEPENDENT OBSERVATIONS

2.1 Motivation

Fusion of data from heterogeneous sources of information, observing a certain phenomenon, has been shown to improve the performance of several inference tasks. Sensors are said to be heterogeneous if their respective observation models cannot be described by the same probability density function (PDF) [51]. Naturally, an information fusion system comprising multi-modal sensors satisfies this definition. However, sensors of the same modality too can be heterogeneous, in the sense defined here, as they may span varied deployment and manufacturing conditions.

In this chapter, we consider centralized detection in the Neyman-Pearson (NP) framework with heterogeneous dependent observations for two cases. The first one is when the dependence is non-stationary where the non-stationarity is assumed to manifest itself as time-varying spatial dependence across the sensors. This is a plausible situation, especially in multi-modal deployments: based on the physics governing the individual modalities, transient phenomena

may affect one modality more drastically than the other. This would, therefore, cause the intermodal dependence to fluctuate, but leave the marginal models relatively invariant within the same observation window. In other words, for reasonably short observation windows, the signal from a single modality can be modeled as a quasi-stationary process, an approach that has been used extensively in spectral analysis and statistical signal processing [36, 70]; modeling cross-sensor dependence, on the other hand, would require a more considered approach. The other case is when the true dependence is beyond the description of any single copula. Under such circumstances, using a single copula to represent the true dependence structure will introduce model mismatch error and lead to suboptimal performance [50]. Thus, a mixture of copulas model which is able to approach the true dependence in multiple directions can be used . A mixture of copulas is usually more flexible compared with single copula models due to an increase in the number of parameters, thus resulting in more degree of freedom and a better performance in describing the dependence among observations.

2.2 Problem Formulation

Consider a scene or phenomenon being monitored by a sensor suite, consisting of N sensors. The nth sensor, $n = 1, 2, \ldots, N$, makes a set of L measurements, $x_{nl}, l = 1 \ldots, L$. These measurements may represent a time series (with l being the time index), spectral coefficients (with l being the frequency index), or some other feature vector. The vector $\mathbf{x}_l$ denotes the l-th measurements at all the sensors, i.e., $\mathbf{x}_l = [x_{1l}, \ldots, x_{Nl}]^T$. The collective measurements, $\mathbf{x} = [\mathbf{x}_1, \ldots, \mathbf{x}_L]$, are received at a processing unit or Fusion Center (FC). Based on the *joint* characteristics of $\mathbf{x}$, the FC decides whether a phenomenon is present or absent in the region

of interest and, thus, solves the following hypothesis testing problem:

$$\begin{aligned} H_0 : f(\mathbf{x}\,|H_0) &= \prod_{l=1}^{L} f(\mathbf{x}_l|H_0) \\ H_1 : f(\mathbf{x}\,|H_1) &= \prod_{l=1}^{L} f(\mathbf{x}_l|H_1), \end{aligned} \tag{2.1}$$

where H_0 is the null hypothesis that the background process is observed, and H_1 is the alternative, i.e., the phenomenon of interest is observed. The PDFs under the null and alternative hypotheses are, respectively, denoted as $f(\cdot|H_0)$ and $f(\cdot|H_1)$. In taking the product over all l in (2.1), we assume that for a given sensor, signals are independent over the index l, e.g., over time. However, in general,

$$f(\mathbf{x}_l\,|H_i) \neq \prod_{n=1}^{N} f_n(x_{nl}|H_i), \quad i = 0, 1 \tag{2.2}$$

where $f_n(x_n|H_0)$ and $f_n(x_n|H_1)$, respectively denote the PDFs of sensor n's observations under hypotheses H_0 and H_1. This formulation, therefore, asserts that since the sensors are observing the same phenomenon, at any given instant, sensor measurements need not be independent spatially (across sensors).

Using Sklar's theorem (Section 1.1.1, Theorem 1.1), the joint densities in (2.1) can be expressed in terms of the copula densities, c_0 and c_1, respectively under H_1 and H_0, as,

$$f(\mathbf{x}\,|H_i) = \prod_{l=1}^{L} \left[\left(\prod_{n=1}^{N} f_n(x_{nl}|\boldsymbol{\theta}_{in}, H_i) \right) c_i(u^i_{1l}(\boldsymbol{\theta}_{i1}), \ldots, u^i_{Nl}(\boldsymbol{\theta}_{iN})|\boldsymbol{\phi}_i) \right], i = 0, 1 \tag{2.3}$$

The copula arguments are the probability integral transforms (PIT) of x_{nl} under hypothesis H_i, i.e., for sensor n and measurement l,

$$u^i_{nl}(\boldsymbol{\theta}_{in}) = F_n(x_{nl}|\boldsymbol{\theta}_{in}, H_i) \quad i = 0, 1. \tag{2.4}$$

The quantities $\{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1\}$ and $\{\boldsymbol{\phi}_0, \boldsymbol{\phi}_1\}$ in (2.3) are, respectively, the marginal density parameters and copula parameters under $\{H_0, H_1\}$. When these parameters are known, the likelihood ratio test (LRT) is the optimal test. Equivalently, we can compare the log-likelihood ratio (LLR) to a threshold η,

$$T_{\mathsf{LR}}(\mathbf{x}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{2.5}$$

where,

$$\begin{aligned} T_{\mathsf{LR}}(\mathbf{x}) &= \log \frac{f(\mathbf{x}\,|H_1)}{f(\mathbf{x}\,|H_0)} \\ &= \sum_{l=1}^{L}\sum_{n=1}^{N} \log \frac{f_n(x_{nl}|\boldsymbol{\theta}_{1n}, H_1)}{f_n(x_{nl}|\boldsymbol{\theta}_{0n}, H_0)} \\ &\quad + \sum_{l=1}^{L} \log \frac{c_1(u^1_{1l}(\boldsymbol{\theta}_{11}), \ldots, u^1_{Nl}(\boldsymbol{\theta}_{1N})|\boldsymbol{\phi}_1)}{c_0(u^0_{1l}(\boldsymbol{\theta}_{01}), \ldots, u^0_{Nl}(\boldsymbol{\theta}_{0N})|\boldsymbol{\phi}_0)} \end{aligned} \tag{2.6}$$

These parameters $\{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1\}$ and $\{\boldsymbol{\phi}_0, \boldsymbol{\phi}_1\}$ are typically unknown and have to be estimated. Using maximum likelihood (ML) estimates in place of the true parameter values, the test becomes a generalized likelihood ratio test (GLRT) in the Neyman-Pearson framework. From (2.3) and (2.4), it is seen that the copula density is also a function of the marginal parameter, $\boldsymbol{\theta}_{in}$, through the PIT. Thus, ideally, ML estimation of the parameters would require simultaneous maximization of the joint likelihood function over both, the marginal and copula parameters. This is, however, difficult and a consistent two-step estimation procedure is commonly used in the copula literature [108]. The two-step maximum likelihood (TSML) procedure first maximizes the individual marginal likelihoods over each $\boldsymbol{\theta}_{in}$:

$$\widehat{\boldsymbol{\theta}}_{in} = \arg\max_{\boldsymbol{\theta}_{in}} \sum_{l=1}^{L} \log f_n(x_{nl}|\boldsymbol{\theta}_{in}, H_i) \tag{2.7}$$

The second step of TSML substitutes $\boldsymbol{\theta}_{in} = \widehat{\boldsymbol{\theta}}_{in}$ in (2.4); the copula likelihood, thus obtained,

is then maximized over $\boldsymbol{\phi}_i$, i.e.,

$$\widehat{\boldsymbol{\phi}}_i = \arg\max_{\boldsymbol{\phi}_i} \sum_{l=1}^{L} \log c_i(\hat{u}^i_{1l}, \ldots, \hat{u}^i_{Nl}|\boldsymbol{\phi}_i). \tag{2.8}$$

where $\hat{u}^i_{nl} = u^i_{nl}(\widehat{\boldsymbol{\theta}}_{in})$. The GLRT then can be expressed as,

$$T_{\mathsf{GLR}}(\mathbf{x}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{2.9}$$

where,

$$\begin{aligned} T_{\mathsf{GLR}}(\mathbf{x}) = &\sum_{l=1}^{L}\sum_{n=1}^{N} \log \frac{f_n(x_{nl}|\widehat{\boldsymbol{\theta}}_{1n}, H_1)}{f_n(x_{nl}|\widehat{\boldsymbol{\theta}}_{0n}, H_0)} \\ &+ \sum_{l=1}^{L} \log \frac{c_1(u^1_{1l}(\widehat{\boldsymbol{\theta}}_{11}), \ldots, u^1_{Nl}(\widehat{\boldsymbol{\theta}}_{1N})|\widehat{\boldsymbol{\phi}}_1)}{c_0(u^0_{1l}(\widehat{\boldsymbol{\theta}}_{01}), \ldots, u^0_{Nl}(\widehat{\boldsymbol{\theta}}_{0N})|\widehat{\boldsymbol{\phi}}_0)} \end{aligned} \tag{2.10}$$

Alternatively, for the bivariate case ($N = 2$), we can also use the sample estimate of Kendall's τ, defined in (1.9), to estimate $\boldsymbol{\phi}_i$. Noting that the relation in (1.9) is invertible, we rewrite the function relationship between τ and $\boldsymbol{\phi}_i$ in (1.7), in terms of a function g_i so that, $\boldsymbol{\phi}_i = g_i^{-1}(\tau)$. The resultant estimate of $\boldsymbol{\phi}_i$ is given by

$$\widetilde{\boldsymbol{\phi}}_i = g_i^{-1}(\hat{\tau}). \tag{2.11}$$

In multivariate copulas, such as Gaussian copula, the dependence parameter is a matrix where each element is the correlation coefficient of two random variables and can be associated with the corresponding pairwise τ. In other forms of multivariate copulas, such as the ones constructed through a vine [65, 97], (conditional) bivariate copulas are the basic elements of the structure, whose associated parameters are directly related to the corresponding (conditional) pairwise τ.

Further, since $\hat{\tau}$ is a consistent estimator of τ [42], $\widetilde{\boldsymbol{\phi}}_i \to \boldsymbol{\phi}_i$ as $L \to \infty$. For finite L, using $\widetilde{\boldsymbol{\phi}}_i$ instead of $\widehat{\boldsymbol{\phi}}_i$, in (2.10), results in a sub-optimal test, but a simpler estimation procedure.

2.3 Detection under Non-stationary Dependence

We motivated the need to consider non-stationary dependence earlier in this chapter. While the preceding section assumes that the family of copulas, c_0 and c_1, are known, a formulation with non-stationary dependence has to necessarily drop that assumption. In the following discussion, we assume that the background model can be predetermined to some degree: the family of the marginals is known and c_0 is known. The more general case of unknown c_0 is considered by Iyengar et al. [51], but signal detection for such a scheme need to be implemented under a training-testing paradigm. However, non-stationarity notwithstanding, the true underlying copula under H_1, c, is typically not known; this "true copula" is usually abstracted as a single copula, but it may, in fact, be a composite of several copulas interacting in an indeterminate fashion, accounting for the non-stationary nature of observations. Due to these complexities, copula selection is an important part of copula based inference and several copula selection methods have been proposed [51, 97, 103]. Our assumptions are stated more precisely as follows:

1. We assume that $f_n(\cdot|H_0)$, the marginal density families under H_0, are known for each $n = 1, \ldots, N$. The corresponding marginal parameters, $\boldsymbol{\theta}_{0n}$, may be unknown.

2. The H_0 copula family, c_0, is known but $\boldsymbol{\phi}_0$ may be estimated, if needed. This section, however, assumes, without loss of generality, that $c_0 = 1$, i.e., measurements under H_0 are independent across sensors. However, the discussion is valid for any known c_0. The independence under the null hypothesis also allows us to simplify our notation; we do not explicitly notate for H_1 with respect to the copula function. Therefore, we set

$$c_1(\cdot) \equiv c(\cdot)$$
$$u_{nl}^1(\boldsymbol{\theta}_{in}) \equiv u_{nl}$$
$$\boldsymbol{\phi}_1 \equiv \boldsymbol{\phi}$$

3. The copula under the alternative, c_1, is not known *a priori*. The "best" copula, in the sense of maximum likelihood, is selected from a predefined library of copulas, $\mathcal{C} = \{c_m : m = 1, \ldots, M\}$.

Based on these assumptions, we discuss three detection scenarios: detection with known parameters, detection with unknown parameters, and detection with unknown marginals under H_1 and unknown copula parameters.

2.3.1 Detection with Known Parameters

For some applications, it may be feasible to determine, *a priori*, the value of the copula parameter $\boldsymbol{\phi}_m$ for each $c_m \in \mathcal{C}$. The actual selection of the copula may be done online. For this case the test-statistic is formulated as a modification of (2.6),

$$\begin{aligned} T_{\mathsf{LR}}(\mathbf{x}) &= \log \frac{f(\mathbf{x}\,|H_1)}{f(\mathbf{x}\,|H_0)} \\ &= \sum_{l=1}^{L}\sum_{n=1}^{N} \log \frac{f_n(x_{nl}|\boldsymbol{\theta}_{1n}, H_1)}{f_n(x_{nl}|\boldsymbol{\theta}_{0n}, H_0)} \\ &\quad + \sum_{l=1}^{L} \log c_l^*(u_{1l}, \ldots, u_{Nl}|\boldsymbol{\phi}_l^*), \end{aligned} \tag{2.12}$$

where for each l the maximum copula likelihood is selected from the library $\mathcal{C}$, i.e.,

$$c_l^* = \arg\max_{c_m \in \mathcal{C}} c_m(u_{1l}, \ldots, u_{Nl}|\boldsymbol{\phi}_l^*) \tag{2.13}$$

The key difference here is that previous work has proposed the scheme of selecting a single copula for the entire observation window [51, 97], i.e., choose a single copula $c_{\mathbf{L}}^*$

$$c_{\mathbf{L}}^* = \arg\max_{c_m \in \mathcal{C}} \sum_{l=1}^{L} \log c_m(u_{1l}, \ldots, u_{Nl}|\boldsymbol{\phi}_{\mathbf{L}}^*) \tag{2.14}$$

for all $l = 1, \ldots, L$. On the other hand, we select the best copula for *each* l adapting to

potentially changing dependence structure.

In problems where the distribution of measurements under one of the hypotheses (based on our assumptions, H_0) is known, it has been shown that [37] the loss in detection power defined as

$$\Delta_{\text{loss}} = \int P_D(\eta) - \hat{P}_D(\eta) d\eta \tag{2.15}$$

when $f_{\mathbf{X}}(\cdot|H_1)$ is misspecified as $\hat{f}(\cdot|H_1)$, is equal to the Kullback-Leibler (KL) divergence $D(f_{\mathbf{X}}(\cdot|H_1)||\hat{f}(\cdot|H_1))$. We now prove that selecting the best copula for each l, as opposed to a single best copula for all L, leads to a smaller KL divergence from the single true copula, c and thus a better performance in terms of loss in detection power.

Proposition 2.1. *Let* $\mathbf{X} \sim f(\mathbf{x}\,|H_1), \mathbf{X} \in \mathbb{R}^{L\times N}$*, where,*

$$f_{\mathbf{X}}(\mathbf{x}\,|H_1) = \prod_{l=1}^{L} \left[\left(\prod_{n=1}^{N} f_n(x_{nl}|\boldsymbol{\theta}_{1n}, H_1) \right) \mathsf{c}(u_{1l}, \ldots, u_{Nl}|\boldsymbol{\phi}) \right] \tag{2.16}$$

where c *is the true copula. For the copula library,* $\mathcal{C} = \{c_m : m = 1, \ldots, M\}$*, and selection schemes* (2.13) *and* (2.14)*,*

$$D(f_{\mathbf{X}}||f_{c_l^*}) \leq D(f_{\mathbf{X}}||f_{c_{\mathbf{L}}^*}), \tag{2.17}$$

where $f_{c_l^*}$ *and* $f_{c_{\mathbf{L}}^*}$ *are the joint densities for* $\mathbf{X}$ *under* H_1 *using* (2.13) *and* (2.14) *respectively.*

Proof. Consider the case $M = 2$. Choosing c_1 over c_2 when $c_1(u_{1l}, \ldots, u_{Nl}) \geq c_2(u_{1l}, \ldots, u_{Nl})$ is equivalent to the decision rule when copula selection is posed as a decision problem with equally likely copulas. Let Ω represent the sample space. Let $\Omega_m \subset \Omega$ represent the decision region for $\mathbf{x}$ for which c_m $(m = 1, 2)$ is chosen, so that $\Omega_1 \cup \Omega_2 = \Omega$ and $\Omega_1 \cap \Omega_2 = \emptyset$. Denote the product of marginals as $f_p(\mathbf{x}_l)$, i.e.,

$$f_p(\mathbf{x}_l) = \prod_{n=1}^{N} f_n(x_{nl}|\boldsymbol{\theta}_{1n}, H_1) \tag{2.18}$$

Also, define the following sets:

$$\mathcal{J}_1 = \{l : \mathbf{x}_l \in \Omega_1\} \text{ and } \mathcal{J}_2 = \{l : \mathbf{x}_l \in \Omega_2\}.$$

Then,

$$\begin{aligned} D(f_{\mathbf{X}}||f_{c_l^*}) &= \int \log \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{c_l^*}(\mathbf{x})} dF_{\mathbf{X}} \\ &= \int \log \frac{\prod_{l=1}^{L} f_p(\mathbf{x}_l)\mathsf{c}(u_{1l}, \ldots, u_{Nl}|\boldsymbol{\phi})}{\prod_{l=1}^{L} f_p(\mathbf{x}_l) \prod_{\mathcal{J}_1} c_1(\cdot|\boldsymbol{\phi}_1) \prod_{\mathcal{J}_2} c_2(\cdot|\boldsymbol{\phi}_2)} dF_{\mathbf{X}} \\ &= \int \sum_{l=1}^{L} \log \mathsf{c}(\cdot) dF_{\mathbf{X}} \\ &\quad - \int \left[\sum_{\mathcal{J}_1} \log c_1(\cdot) + \sum_{\mathcal{J}_2} \log c_2(\cdot) \right] dF_{\mathbf{X}} \end{aligned} \tag{2.19}$$

The selection criterion in (2.13), implies that, for the set $\mathcal{J}_2$, $c_2 \geq c_1$. Therefore,

$$\begin{aligned} \sum_{\mathcal{J}_1} \log c_1(\cdot) + \sum_{\mathcal{J}_2} \log c_2(\cdot) &\geq \sum_{\mathcal{J}_1} \log c_1(\cdot) + \sum_{\mathcal{J}_2} \log c_1(\cdot) \\ &= \sum_{l=1}^{L} \log c_1(\cdot), \end{aligned} \tag{2.20}$$

and in a similar manner,

$$\sum_{\mathcal{J}_1} \log c_1(\cdot) + \sum_{\mathcal{J}_2} \log c_2(\cdot) \geq \sum_{l=1}^{L} \log c_2(\cdot) \tag{2.21}$$

Therefore, depending on whether c_1 or c_2 was chosen using (2.14), we can substitute either of the inequalities in (2.20) and (2.21) in (2.19) to get,

$$D(f_{\mathbf{X}}||f_{c_l^*}) \leq D(f_{\mathbf{X}}||f_{c_{\mathbf{L}}^*})$$

This proves the case for $M = 2$. For $M > 2$ we can successively partition Ω_2 and arrive at a similar result by repeating the above steps. It should be noted that, in the proof, we assume the true dependence to be c for all time instants l. Since the true copula appears only in the first term of (2.19), which is irrelevant to the copula selection rule, the proof holds for any type of true dependence (time invariant, or time varying). □

Proposition 2.1 implies that a detector using the selection scheme proposed in (2.13) will suffer a lower loss in detection performance due to copula misspecification [51].

2.3.2 Detection with Unknown Parameters

With unknown parameters, the statistic in (2.10) for the composite hypothesis testing problem can be rewritten as,

$$\begin{aligned} T_{\mathsf{GLR}}(\mathbf{x}) = &\sum_{l=1}^{L}\sum_{n=1}^{N} \log \frac{f_n(x_{nl}|\widehat{\boldsymbol{\theta}}_{1n}, H_1)}{f_n(x_{nl}|\widehat{\boldsymbol{\theta}}_{0n}, H_0)} \\ &+ \sum_{l=1}^{L} \log c_l^*(u_{1l}(\widehat{\boldsymbol{\theta}}_{11}), \ldots, u_{Nl}(\widehat{\boldsymbol{\theta}}_{1N})|\widehat{\boldsymbol{\phi}}_l^*), \end{aligned} \tag{2.22}$$

where the TSML procedure has been used to obtain estimates of marginal and copula parameters. The copula parameters $\widehat{\boldsymbol{\phi}}_m$ are estimated over L, for each $c_m \in \mathcal{C}$, so that

$$\mathcal{C} = \{c_m(\cdot|\widehat{\boldsymbol{\phi}}_m(L)) : m = 1, \ldots, M\} \tag{2.23}$$

$$c_l^* = \arg\max_{c_m \in \mathcal{C}} \log c_m(u_{1l}(\widehat{\boldsymbol{\theta}}_{11}), \ldots, u_{Nl}(\widehat{\boldsymbol{\theta}}_{1N})|\widehat{\boldsymbol{\phi}}_m(L)) \tag{2.24}$$

$$\widehat{\boldsymbol{\phi}}_l^* = \arg c_l^* \tag{2.25}$$

While this selection method is motivated by the implications of Proposition 2.1 for the simple hypothesis case, a similar result may not be stated for the composite test. This is because ML estimation requires that all L samples be drawn from the same population; this need not hold

true for copula selection from $\mathcal{C}$ with unknown parameters. The copula parameters can also be estimated using Kendall's $\hat{\tau}$. The test-statistic is then,

$$\begin{aligned} T_{\hat{\tau}}(\mathbf{x}) = &\sum_{l=1}^{L}\sum_{n=1}^{N} \log \frac{f_n(x_{nl}|\widehat{\boldsymbol{\theta}}_{1n}, H_1)}{f_n(x_{nl}|\widehat{\boldsymbol{\theta}}_{0n}, H_0)} \\ &+ \sum_{l=1}^{L} \log c_l^*(u_{1l}(\widehat{\boldsymbol{\theta}}_{11}), \ldots, u_{Nl}(\widehat{\boldsymbol{\theta}}_{1N})|\widetilde{\boldsymbol{\phi}}_l^*), \end{aligned} \tag{2.26}$$

where $\widetilde{\boldsymbol{\phi}}_l^*$ is the estimate of $\boldsymbol{\phi}_l^*$ based on $\hat{\tau}$. Correspondingly,

$$\mathcal{C} = \{c_m(\cdot|\boldsymbol{\phi}_m(\hat{\tau})) : m = 1, \ldots, M\} \tag{2.27}$$

$$c_l^* = \arg\max_{c_m \in \mathcal{C}} \log c_m(u_{1l}(\widehat{\boldsymbol{\theta}}_{11}), \ldots, u_{Nl}(\widehat{\boldsymbol{\theta}}_{1N})|\boldsymbol{\phi}_m(\hat{\tau})) \tag{2.28}$$

$$\widetilde{\boldsymbol{\phi}}_l^* = \arg c_l^* \tag{2.29}$$

2.3.3 Detection with Unknown Marginals and Unknown Copula Parameters

In many applications, establishing a model under H_1 is not feasible. In that case, $f_n(\cdot|H_1)$ is determined non-parametrically and u_{ij} is obtained using the empirical probability integral transform (EPIT). The test statistic is, therefore, expressed as,

$$\begin{aligned} T_{\mathsf{NPM}}(\mathbf{x}) = &\sum_{l=1}^{L}\sum_{n=1}^{N} \log \frac{\hat{f}_n(x_{nl}|H_1)}{f_n(x_{nl}|\widehat{\boldsymbol{\theta}}_{0i}, H_0)} \\ &+ \sum_{l=1}^{L} \log c_l^*(\hat{u}_{1l}, \ldots, \hat{u}_{Nl}|\widehat{\boldsymbol{\phi}}_l^*), \end{aligned} \tag{2.30}$$

where c_l^* and associated parameters are selected as indicated in (2.23), (2.24) and (2.25). The uniform random variables in the copula density are evaluated using EPIT,

$$\hat{F}_n(\cdot|H_1) = \frac{1}{L}\sum_{l=1}^{L+1} \mathbb{1}_{x_{nl}<\{\cdot\}} \tag{2.31}$$

$$\hat{u}_{nl} = \hat{F}_n(x_{nl}|H_1) \tag{2.32}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function.

The marginal model under H_1 is determined through a kernel density estimation procedure. Kernel density estimators [119] provide a smoothed estimate, $\hat{f}_n(x_{nl}|H_1)$, of the true density. Choosing the correct bandwidth for kernel density estimation is important for an accurate estimate. The kernel bandwidth is chosen using leave-one-out cross-validation. The selected bandwidth, h^*, is the minimizer of the cross-validation estimator of risk, $\hat{J}$, for a kernel, K. The risk estimator may be easily computed using the approximation [119, p. 136],

$$\begin{aligned}\hat{J}(h) = \frac{1}{hN^2}\sum_p\sum_q K^*\left(\frac{X_p - X_q}{h}\right) \\ + \frac{2}{Nh}K(0) + \mathcal{O}\left(\frac{1}{N^2}\right),\end{aligned} \tag{2.33}$$

where,

$$\begin{aligned}K^*(x) &= K^{(2)}(x) - 2K(x) \\ K^{(2)}(z) &= \int K(z-y)K(y)dy.\end{aligned}$$

The Gaussian kernel is used in this work, so that $K(x) = \mathcal{N}(x;0,1)$ and $K^{(2)}(z) = \mathcal{N}(z;0,2)$. Therefore,

$$h^* = \arg\min_h \hat{J}(h)$$

2.4 Detection Using Mixture of Copulas

In the previous section, for each sample we chose one copula as the "best" one from the copula library $\mathcal{C}$. We actually made the assumption that at each time instant there is an unknown true copula from which the data is generated, thus a hard decision on the best copula is made at each time instant as in a multiple hypothesis testing problem. However, in this section, we assume that every sample is generated by c_m with probability $\pi_m \in [0,1], \forall m = 1, \ldots, M$, i.e.,

$$c(\cdot) = \sum_{m=1}^{M} \pi_m c_m(\cdot|\phi_m), \tag{2.34}$$

where $\{c_m(\cdot|\phi_m)\}$ are the copula components from library $\mathcal{C}$, $\{\pi_m\}_{m=1}^{M}$ are the corresponding weights satisfying $\sum_{m=1}^{M} \pi_m = 1$. It has been proved in [75] that (2.34) is also a copula. The mixture of copulas model provide a more flexible approach to model dependence and can serve as a better descriptor of dependence than a single copula.

The assumptions 1, 2 and 3 that we made in Section 2.3 are still assumed to hold here. According to Sklar's theorem and the mixed copula in (2.34), the distribution under H_1 can be written as

$$f(\mathbf{x}) = \prod_{l=1}^{L} \left[\left(\prod_{n=1}^{N} f_n(x_{nl}|\boldsymbol{\theta}_n) \right) \times \sum_{m=1}^{M} \pi_m c_m(u_{1l}(\boldsymbol{\theta}_1), \ldots, u_{Nl}(\boldsymbol{\theta}_N)|\phi_m) \right]. \tag{2.35}$$

Based on the TSML procedure, marginal parameters can be first obtained by (2.7). Then, after substituting $\theta_n = \hat{\theta}_n, n = 1, \ldots, N$, the copula likelihood function is obtained and maximized over unknown copula parameters $\boldsymbol{\phi} = (\boldsymbol{\phi_1}, \ldots, \boldsymbol{\phi_M})$ and weights $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_M)$.

$$\widehat{\boldsymbol{\pi}}, \widehat{\boldsymbol{\phi}} = \arg\max_{\boldsymbol{\pi}, \boldsymbol{\phi}} \sum_{l=1}^{L} \log \sum_{m=1}^{M} \pi_m c_m(\hat{u}_{1l}, \ldots, \hat{u}_{Nl}|\boldsymbol{\phi_m}). \tag{2.36}$$

Then the GLRT can be written as:

$$T_{\mathsf{GLR}}(\mathbf{x}\,|\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\pi}}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{2.37}$$

where

$$\begin{aligned} T_{\mathsf{GLR}}(\mathbf{x}) = &\sum_{l=1}^{L}\sum_{n=1}^{N} \log \frac{f_n(x_{nl}|\widehat{\boldsymbol{\theta}}_{1n}, H_1)}{f_n(x_{nl}|\widehat{\boldsymbol{\theta}}_{0n}, H_0)} \\ &+ \sum_{l=1}^{L} \log \sum_{m=1}^{M} \hat{\pi}_m c_m(\hat{u}_{1l}, \ldots, \hat{u}_{Nl}|\widehat{\boldsymbol{\phi}}_m), \end{aligned} \tag{2.38}$$

The superiority of using a mixture copula model, compared with using a single copula model is given in Proposition 2.2:

Proposition 2.2. *Let* $\mathbf{X} \sim f(\mathbf{x}), \mathbf{X} \in \mathbb{R}^{1\times N}$*, where,*

$$f_{\mathbf{X}}(\mathbf{x}) = \left[\left(\prod_{n=1}^{N} f_n(x_n|\boldsymbol{\theta}_n)\right) \mathsf{c}(u_1, \ldots, u_N|\boldsymbol{\phi})\right] \tag{2.39}$$

where c *is the true copula. As* $L \to \infty$

$$D(f_{\mathbf{X}}||f_{c^*_{\boldsymbol{M}}}) \leq D(f_{\mathbf{X}}||f_{c^*}), \tag{2.40}$$

where $f_{c^*_{\boldsymbol{M}}}$ *and* f_{c^*} *are the joint densities for* $\mathbf{X}$ *under* H_1 *using mixture of copulas and single copula respectively.*

Proof. Consider the case $M = 2$. $f_p(\mathbf{x})$ is defined in (2.18). Then,

$$\begin{aligned} D(f_{\mathbf{X}}||f_{c^*_{\boldsymbol{M}}}) &= \int \log \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{c^*_{\boldsymbol{M}}}(\mathbf{x})} dF_{\mathbf{X}} \\ &= \int \log \frac{f_p(\mathbf{x})\mathsf{c}(u_1, \ldots, u_L|\boldsymbol{\phi})}{f_p(\mathbf{x})[\pi_1 c_1(\cdot|\boldsymbol{\phi}_1) + \pi_2 c_2(\cdot|\boldsymbol{\phi}_2)]} dF_{\mathbf{X}} \\ &= \int \log \mathsf{c}(\cdot) dF_{\mathbf{X}} - \int \log[\pi_1 c_1(\cdot|\boldsymbol{\phi}_1) + \pi_2 c_2(\cdot|\boldsymbol{\phi}_2)] dF_{\mathbf{X}} \end{aligned} \tag{2.41}$$

When L is large enough, we use the sample estimate to approach the true distribution, meaning

that when $L \to \infty$

$$\begin{aligned} &\frac{1}{L}\sum_{l=1}^{L}\log[\pi_1 c_1(\cdot|\boldsymbol{\phi}_1)+\pi_2 c_2(\cdot|\boldsymbol{\phi}_2)] \\ \to &\int \log\left[\pi_1 c_1(\cdot|\boldsymbol{\phi}_1)+\pi_2 c_2(\cdot|\boldsymbol{\phi}_2)\right] dF_{\mathbf{X}} \end{aligned} \tag{2.42}$$

So, the second term in (2.41) can be approximated by the sample estimate in (2.42). Since, a single copula model corresponds to a special case of mixture of copulas model in which all the weights are zeros except for one weight that is equal to one, it can be concluded that

$$\begin{aligned} &\max_{\phi_1,\phi_2,\pi_1,\pi_2} \frac{1}{L}\sum_{l=1}^{L}\log\left[\pi_1 c_1(\cdot|\boldsymbol{\phi}_1)+\pi_2 c_2(\cdot|\boldsymbol{\phi}_2)\right] \quad \geq \\ &\max_{\phi_1,\phi_2,\pi_m=1,m=1,2} \frac{1}{L}\sum_{l=1}^{L}\log\left[\pi_1 c_1(\cdot|\boldsymbol{\phi}_1)+\pi_2 c_2(\cdot|\boldsymbol{\phi}_2)\right] \end{aligned} \tag{2.43}$$

The right hand side of (2.43) correspond to the single copula model. Thus, when $L \to \infty$, $D(f_{\mathbf{X}}||f_{c^*_{\boldsymbol{M}}}) \leq D(f_{\mathbf{X}}||f_{c^*})$ is proved. For $M > 2$ we can successively arrive at a similar result by repeating the above steps. □

Since there is no explicit analytical expression for the maximum likelihood estimator in (2.36). We use the expectation maximization (EM) algorithm to find the parameters iteratively. The EM algorithm consists of two steps: E-step computes and updates the conditional probability that our observations come from each component copula, and M-step maximizes the log likelihood to estimate the parameters of each copula. Let $\mathcal{Y} = \{y_l\}_{l=1,\ldots,L}$ whose value informs us as to which component copula "generated" each data item. We assume that $y_l \in 1,\ldots,M$ for each l, and $y_l = m$ if the lth sample was generated by the mth mixture component. Then the two steps of the EM algorithm that we developed for the mixture of copula model are as follows:

- E-step

$$\begin{aligned} p(y_l|\boldsymbol{u}_l, \Theta^r) &= \frac{\pi^r_{y_l} c_{y_l}(\boldsymbol{u}_l|\phi^r_{y_l})}{c(\boldsymbol{u}_l|\Theta^r)} \\ &= \frac{\pi^r_{y_l} c_{y_l}(\boldsymbol{u}_l|\phi^r_{y_l})}{\sum_{m=1}^{M} \pi^r_m c_m(\boldsymbol{u}_l|\phi^r_m)}, \quad \forall l = 1, \ldots, L \end{aligned} \tag{2.44}$$

- M-step

$$\begin{aligned} \pi^{r+1}_m &= \frac{1}{L} \sum_{l=1}^{L} p(y_l|\boldsymbol{u}_l, \Theta^r) \\ \phi^{r+1}_m &= \max_{\phi_m} \sum_{l=1}^{L} \log c_m(\boldsymbol{u}_l|\phi_m) p(y_l|\boldsymbol{u}_l, \Theta^r) \end{aligned} \tag{2.45}$$

where $\Theta = (\boldsymbol{\phi}, \boldsymbol{\pi})$ is the parameter set and the superscript r represents the fact that the parameters estimated at rth step. These two steps are repeated until the stopping criterion is satisfied. Each iteration is guaranteed to increase the log-likelihood and the algorithm is guaranteed to converge to a local maximum of the likelihood function [13].

2.5 Results and Discussion

In this section, we present results when the copula-based tests, discussed in Section 2.3, are applied to simulated and real data. Our results are presented for a two-sensor case, i.e., $N = 2$. We note, however, that the methods described above apply to more generalized cases as well where $N > 2$, as one can construct a multivariate copula using bivariate components [97].

2.5.1 Nonstationary Copula

We assume normal and beta distributed marginals respectively and consider various cases of copula dependence. The marginals and the respective parameters used are tabulated in Ta-

Table 2.1: Distribution of marginals for simulation experiments

n	H_0	H_1
1	$\mathcal{N}(0,1)$	$\mathcal{N}(0.1,1.1)$
2	Beta(2.0,2.0)	Beta(2.2,2.2)

ble 2.1. For all copula cases we used Kendall's $\tau = 0.2$ to specify the amount of dependence. The copula library contains the Gaussian and Frank copulas, i.e., $\mathcal{C} = \{\text{Gaussian}, \text{Frank}\}$. For all the cases, we compare performances obtained via simulation when performing hypothesis testing with T_{GLR} in (2.22), $T_{\hat{\tau}}$ in (2.26), GLR using single copula selection and the product rule, i.e., independence assumption. The results presented are averaged over 10^4 Monte-Carlo trials with 1000 samples per trial.

In Figures 2.1, 2.2 and 2.3 the receiver operating characteristics (ROC) curves are shown when different copulas were used to generate data. In Figure 2.1, we consider the case in which a single student's t copula is used to generate the data. Note that this represents the case where the true copula is not known and is not included in the copula library. The label non-stationary copula refers to the sample-wise copula selection scheme proposed in this chapter. Figure 2.2 represents the case where half of all x_{nl} were generated with a Gaussian copula and the remaining half consisted of samples generated from the Frank copula. This is, therefore, the case where the copula library contains both copula models that were used to generate data. The case of a single generating copula that is also a member of the library is also considered; Figure 2.3 shows results when all the data are generated using the Frank copula. It can be seen from Figures 2.1, 2.2 and 2.3 that for all three cases described above, our sample-wise selection rule outperforms the single copula rule, which has been theoretically justified in Proposition 2.1. All copula-based fusion rules which take inter-sensor dependence into consideration have better performance compared with the fusion rule under independence assumption.

For all simulation scenarios we observe that the GLRT and the test based on $\hat{\tau}$, using our selection scheme perform comparably. Both outperform the single copula selection method

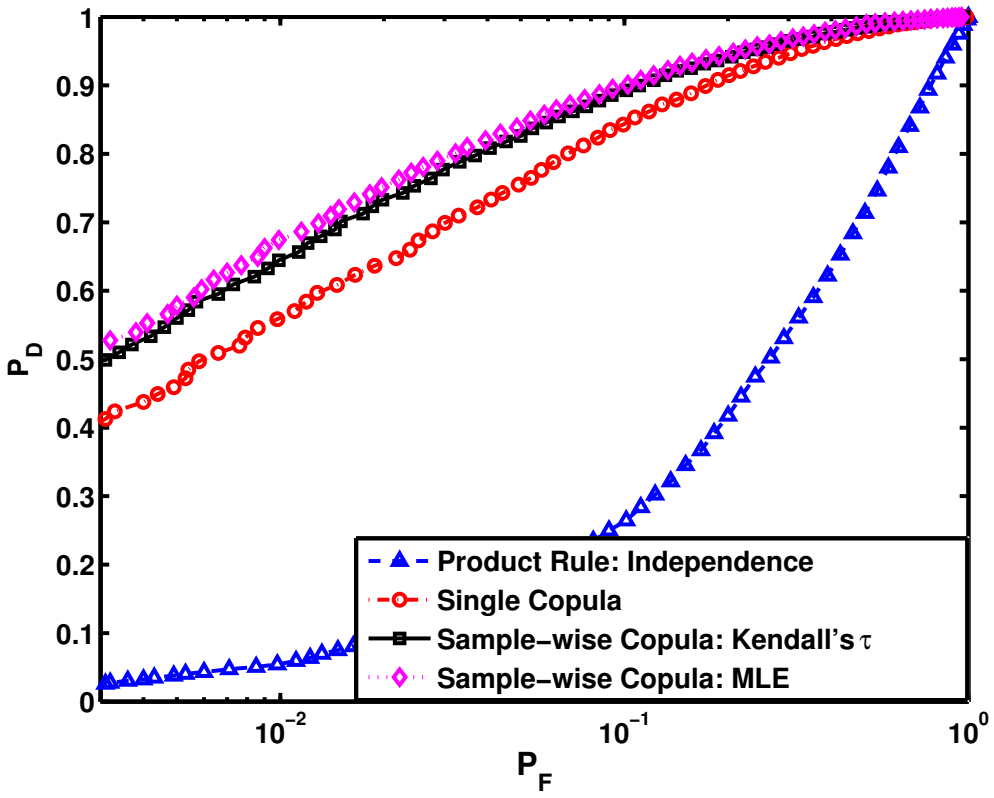

Fig. 2.1: ROCs for H_1 data generated using a t copula

and product rules. We note that these results represent the unknown parameter case (Section 2.3.2), for which we were not able to prove that our method would outperform the single copula selection method. An intuition for why we observe this result is that since $\hat{\tau}$ is consistent, for large L, $\hat{\tau} \rightarrow \tau$. Also, a single value of τ corresponds to different values of $\boldsymbol{\phi}_m = \arg c_m(\cdot), c_m \in \mathcal{C}$. We conjecture that, asymptotically, this is as if the parameter values are known, allowing Proposition 2.1 to be applicable. This implies that while τ controls the *amount* of dependence, which remains unchanged for all L, different copulas represent the *shape* of the dependence between the data from the two sensors.

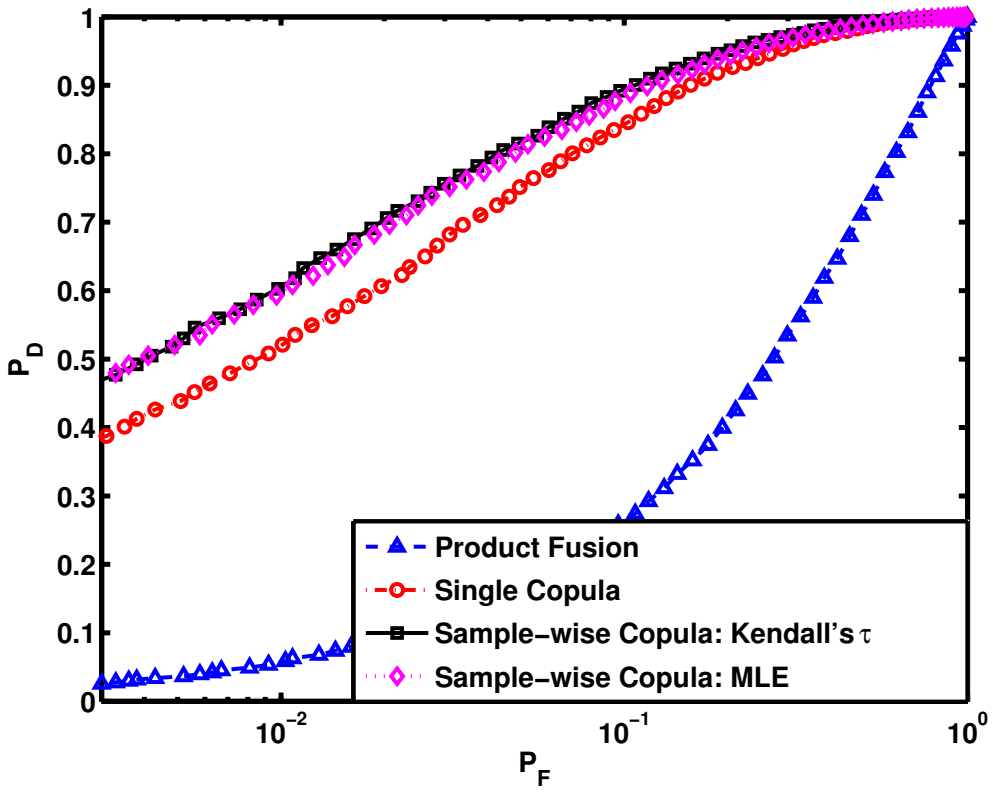

Fig. 2.2: ROCs for H_1 data generated using Frank and Gaussian copulas

2.5.2 Mixture of Copulas

In this subsection, we provide numerical results for the mixture of copula model for a two sensor case, i.e., $N = 2$. We assume normal and beta distributed marginals and consider the mixture of Gaussian and Gumbel, i.e., $\mathcal{C} = \{c_1 = \text{Gaussian}, c_2 = \text{Gumbel}\}$, with different copula parameters. The marginals and the respective parameters used are tabulated in Table 2.1.

We first investigate the performance of the EM algorithm in estimating copula parameters and the corresponding weights, i.e., Θ. When the samples are generated by $c = 0.5c_1(\cdot|\phi_1 = 0.7) + 0.5c_2(\cdot|\phi_2 = 19)$, whose scatter plot is shown in Figure 2.4, The estimated Θ is as follows: $\hat{\phi}_1 = 0.6879, \hat{\pi}_1 = 0.5029; \hat{\phi}_2 = 19.3924, \hat{\pi}_2 = 0.4971$. From Figure 2.4, we can visually distinguish the two different sets of samples generated by different copulas. The estimates are quite close to the true parameters in multiple trials.

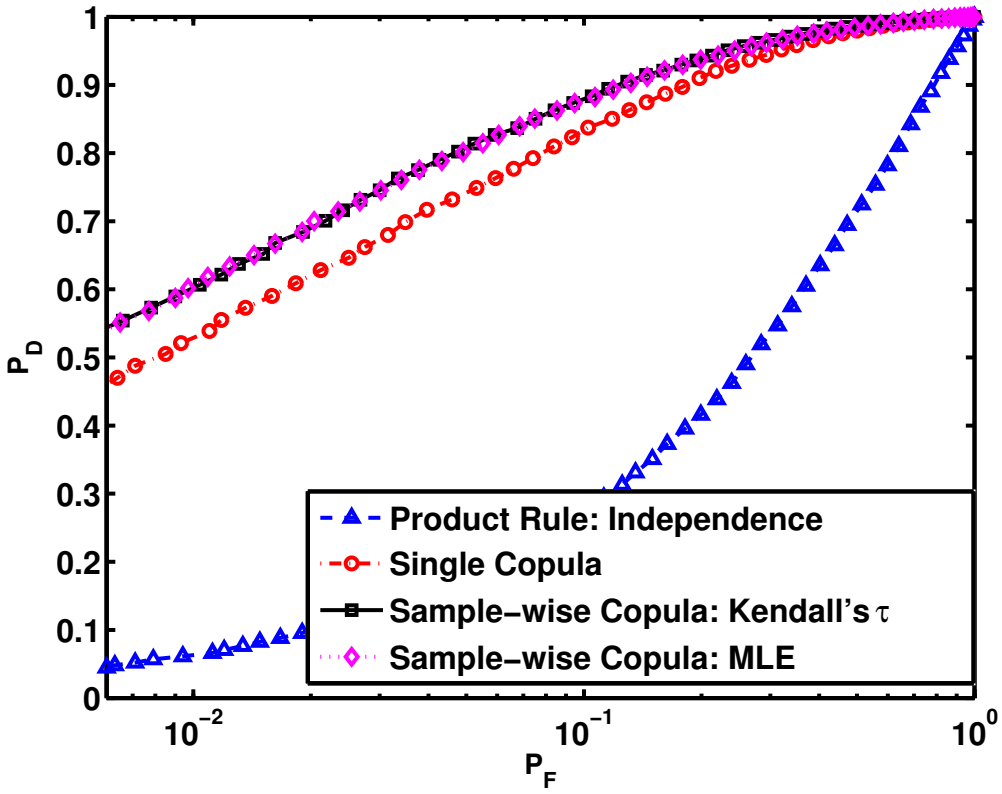

Fig. 2.3: ROCs for H_1 data generated using a Frank copula

When the samples are generated by $c = 0.5c_1(\cdot|\phi_1 = 0.7) + 0.5c_2(\cdot|\phi_2 = 5)$, whose scatter plot is shown in Figure 2.5, The estimated Θ is as follows: $\hat{\phi}_1 = 0.7243, \hat{\pi}_1 = 0.5170; \hat{\phi}_2 = 4.5169, \hat{\pi}_2 = 0.4830$. In Figure 2.5, the two sets of samples are less distinguishable compared with those in Figure 2.4. Also, the estimates vary from trial to trial.

When the samples are generated by $c = 0.5c_1(\cdot|\phi_1 = 0.7) + 0.5c_2(\cdot|\phi_2 = 1.5)$, whose scatter plot is shown in Figure 2.6, The estimated Θ is as follows: $\hat{\phi}_1 = 0.7230, \hat{\pi}_1 = 0.4117; \hat{\phi}_2 = 1.5618, \hat{\pi}_2 = 0.4117$. In Figure 2.6, the two different sets of samples are hardly distinguishable. And the estimation results also show that the estimated π_1 and π_2 do not converge, and their values are different from the true parameters.

In the EM algorithm, we use the convergence of log likelihood as the stopping criterion of the algorithm. Even though, the parameters may not converge to the true values, the log

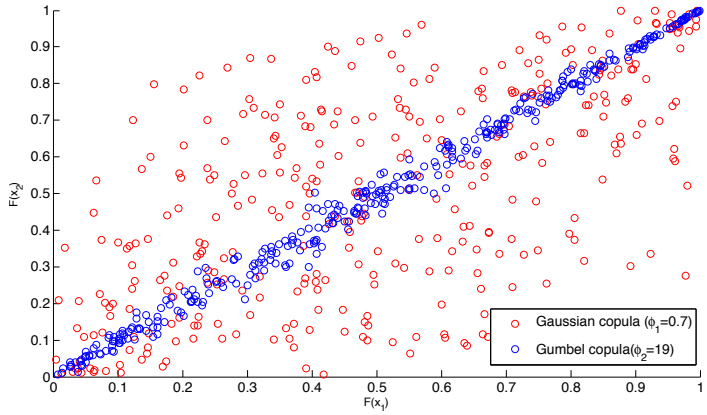

Fig. 2.4: Scatter plot: mixture of Gaussian ($\phi_1 = 0.7, \pi_1 = 0.5$) and Gumbel ($\phi_2 = 19, \pi_2 = 0.5$)

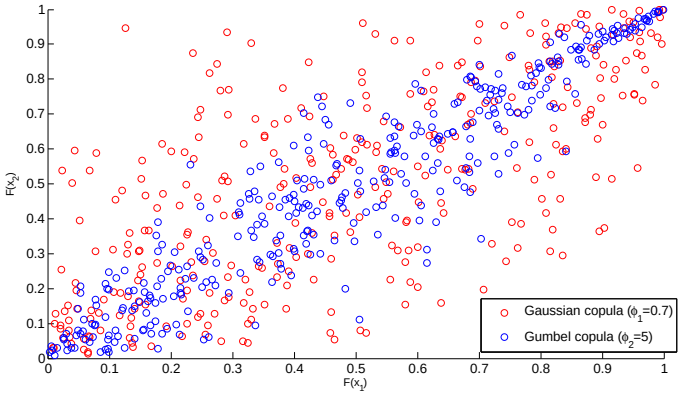

Fig. 2.5: Scatter plot: mixture of Gaussian ($\phi_1 = 0.7, \pi_1 = 0.5$) and Gumbel ($\phi_2 = 5, \pi_2 = 0.5$)

likelihoods converge to the ones that correspond to the true parameters. Since the Likelihood Ratio (LR) does not depend on the precision of these estimated parameters, but on the accuracy of the likelihood score, the detection performance of our approach is promising. We consider the case of data generated using $c = 0.5c_1(\cdot|\phi_1 = 0.7) + 0.5c_2(\cdot|\phi_2 = 1.5)$ as an example to illustrate the detection performance of the mixture of copulas model. Figure 2.7 shows the ROCs for three different detection methods: mixture copula based detection, single copula based detection, and independent model. Using mixture of copulas model to describe the dependence structure under H_1, results in a better detection performance compared with the

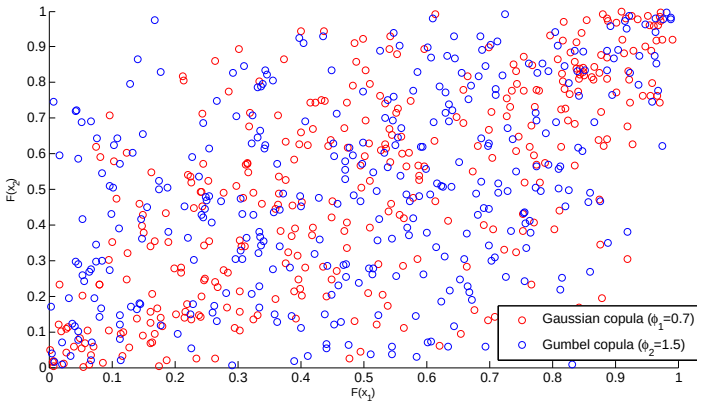

Fig. 2.6: Scatter plot: mixture of Gaussian ($\phi_1 = 0.7, \pi_1 = 0.5$) and Gumbel ($\phi_2 = 1.5, \pi_2 = 0.5$)

single copula model and the independent model.

2.5.3 ARL Footstep Data

In order to test our proposed detection algorithms in the real world, we use the footstep data, made available by the US Army Research Laboratory (ARL), collected at Douglas, AZ to detect personnel activity. The dataset consists of raw observations from several sensors of different modalities that were deployed in an outdoor space to record human and animal activity that is typical in perimeter and border surveillance scenarios. The participants in the data collection exercise walked/ran along a predetermined path with sensors laid out along either side of the path. We consider copula-based seismic-acoustic fusion.

Seismic and acoustic time series for activities representing a single person walking (among other examples) are available in the ARL dataset. A representative dataset from acoustic and seismic sensors is shown in Figure 2.8. Each seismic/acoustic time series contains a leading 60s of background data. We use this as our H_0 data. The data are sampled at 10kHz, and are mean centered and oscillatory in nature.

Before applying the copula-based detector, we first preprocess the data. The time series is split into non-overlapping frames of length $T = 512$. This raw time series data is called $x_{Tn}(t)$

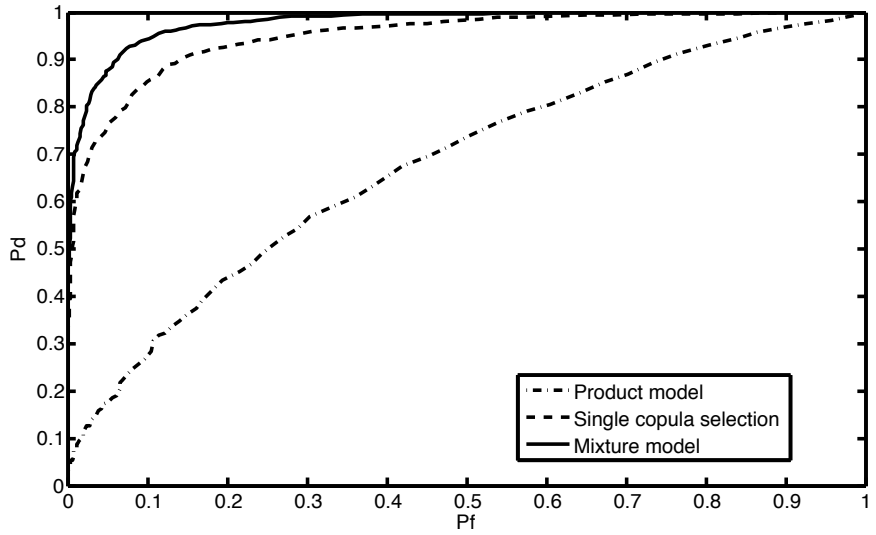

Fig. 2.7: ROCs for H_1 data generated by the mixture of Gaussian and Gumbel.

where $n = 1, 2$ is the sensor index for the acoustic and seismic modalities respectively, and t is the time index. In keeping with Houston's analysis that Fourier spectra for seismic and acoustic footstep data are more informative than time-domain measurements [49], we set

$$x_{nl} = \sqrt{\mathcal{F}\{x_{Tn}(t)\}^2},$$

where $\mathcal{F}$ is the DFT and $l = 1, \ldots, L = 256$ is the frequency index. Our sensor measurements are, therefore, now transformed to the frequency domain and the statistics of $\mathbf{x} = [x_{nl}]$ are used as the input to the detector. The copula library consists of Gaussian, Gumbel and Frank copulas. We have observed that due to the interstitial nature of footstep data, including the independence copula in the library improves the overall detection performance.

For the ARL dataset, the marginal distribution under H_1, i.e., $f_n(\cdot|H_1), n = 1, 2$, is determined non-parametrically and u_{ij} is obtained using the empirical probability integral transform (EPIT). To generate ROCs, we compare the test-statistic to a vector of thresholds. The ROCs thus generated, for detecting H_1 corresponding to one person walking against H_0 corresponding to background noise, are shown in Figure 2.9. We compare the performance of our pro-

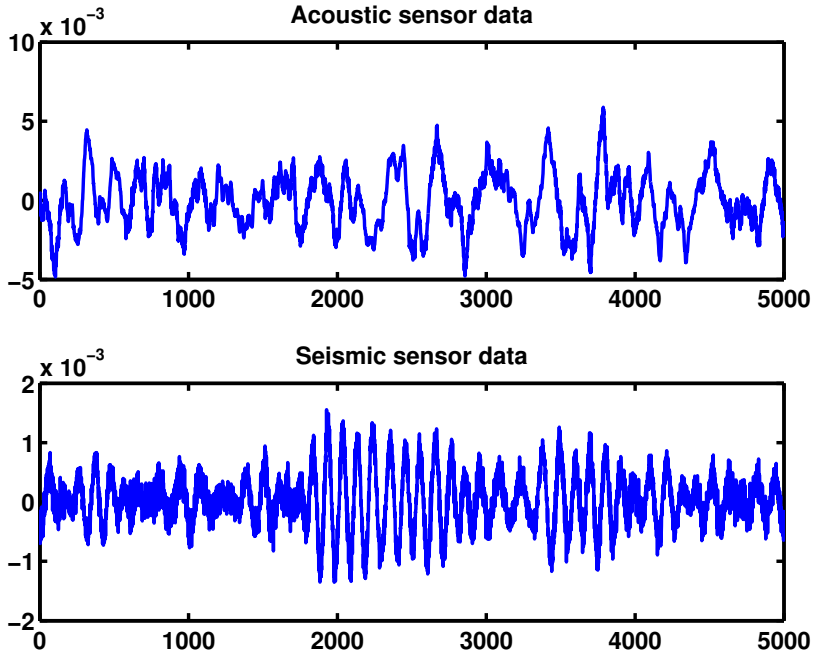

Fig. 2.8: Raw acoustic and seismic sensor data

posed fusion rules based on sample-wise copula selection and mixture of copula model with the performance of other existing fusion rules in the literature. It can be seen from Figure 2.9 that the product rule, which assumes independence under H_1, has the worst performance, since the dependence information among sensor observations is ignored in the fusion rule. The single copula selection rule performs better than the product rule, because the dependence between acoustic and seismic sensor is characterized using a copula density function and taken into consideration in the detector. Our proposed method using the sample-wise copula selection outperforms the fusion rule corresponding to single copula selection. Sample-wise copula selection rule is designed to accommodate for non-stationary data and in this case the non-stationarity is assumed to manifest itself as time-varying spatial dependence across the acoustic sensor and seismic sensors. This is a plausible situation, especially in our outdoor experiment setting where many factors may contribute to the non-stationarity. The superiority of our sample-wise copula selection rule demonstrated through the real footstep data is consistent with Proposition

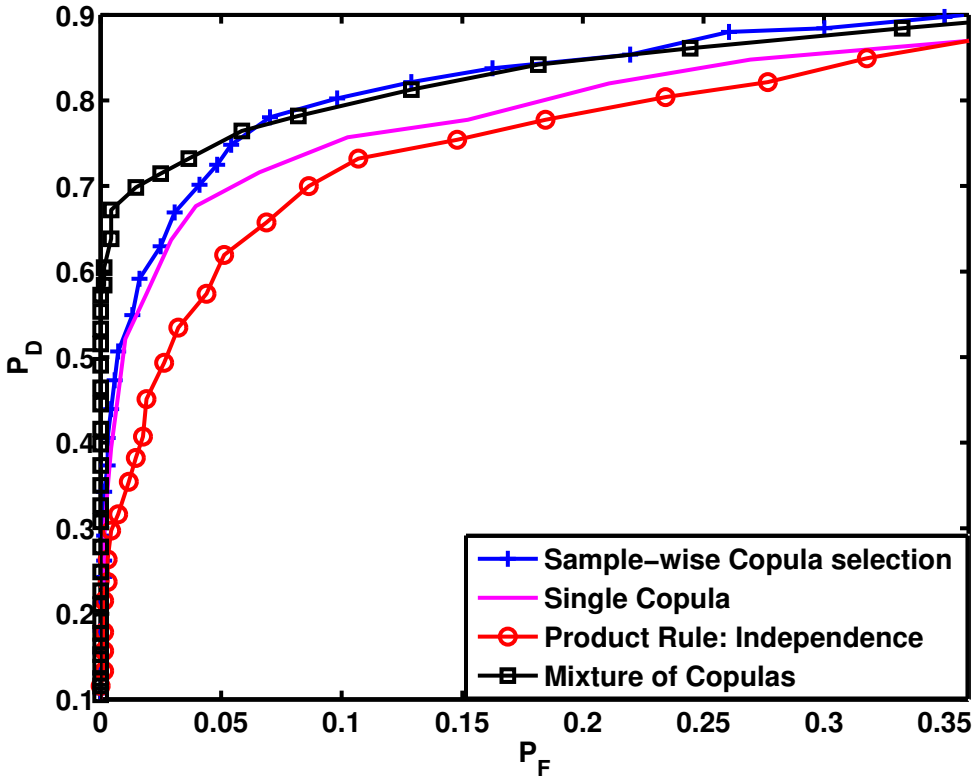

Fig. 2.9: ROCs for the ARL dataset for 1 person vs. background detection.

2.1. The mixture of copula based detector also achieves better performance compared with single copula selection rule. Because the mixture model provides a better description of the real dependence among the two types of sensors and thus results in less performance loss due to copula misspecification.

2.6 Summary

In this chapter, we considered copula-based detection with heterogeneous and dependent sensor observations in an NP framework. According to the nature of the dependence among measurements from different sources, we considered two kinds of situations. When the dependence structure is non-stationary, meaning that the copula which describes the dependence structure

may change over time, we proposed a sample-wise copula selection rule which selects the copula with the maximum likelihood for each sample. The superiority of this method over existing approaches was proved theoretically. When the true dependence follows a mixture model, we designed a detector based on the mixture of copulas. The Expectation Maximization (EM) algorithm was applied for finding the estimates of unknown parameters in the mixture of copulas model. Theoretical proof and simulation results were also provided to demonstrate the performance of our proposed approaches. We also applied our proposed approaches to the processing of real footstep data for personnel detection, where the true nature of dependence among sensor observations are quite complex and unknown. We were able to improve the detection performance using our detectors, compared to other existing approaches.

CHAPTER 3
COPULA-BASED FUSION OF HETEROGENEOUS TIME SERIES

3.1 Motivation

Sensor networks provide information about various aspects of the real world and have become an integral part of various systems used in daily lives. The problem of inferring about events of interest by fusing data from multiple heterogeneous sensors has a wide variety of applications. The inference tasks could consist of detecting an activity of interest or estimating some parameters, such as locations or tracks, which provide actionable intelligence and/or improved situational awareness.

Social media, facilitated by the growth of social networks, provides an easily accessible platform for users to share information and has resulted in the generation of unprecedented amounts of social media data that can be recorded and even monitored (such as, wall posts, clicks etc). This trend is likely to continue with exponentially more content in the future. This massive amount of social media data can be used by the signal processing community for extracting information about unfolding events. This is expected to be beneficial in the military

as well as civilian domains. A number of works have been published regarding the use of social media data for understanding real world phenomena. In particular, social media data have been successfully applied to: prediction of earthquake [89], forecast box office revenues [10], truth discovery [116], prediction of election results [113], stock prediction [15], and automatic crime prediction [118].

For inference with both traditional sensors and social media, with respect to the information content of the signals, information sources exhibit heterogeneity that can arise from a wide variety of causes. As a consequence of heterogeneity, the quality and quantity of information provided by each modality, including human intelligence, vary with each source. For a group of heterogeneous information sources observing the same phenomenon, local observations are statistically dependent, and yet provide different characterizations of the phenomenon under observation. Such diversity in the sensing process not only leads to enhanced inference quality, but also improves fault tolerance capabilities, so that the decision making capability of the system does not become impaired completely due to the failure of one modality. However, an accurate characterization of the intermodal dependence and development of algorithms to jointly process and fuse heterogeneous data are necessary for making reliable system-wide inference.

In this chapter, we develop new techniques for inference, using data from diverse information sources, including both social media and sensor networks. Inference in traditional heterogeneous sensor networks has been investigated in [52, 96, 100] using copula theory, but assuming that the observations are temporally independent. In the networks of sensors and social media, temporal dependence may exist, along with intermodal dependence. We build a copula-based characterization of multivariate time series, in which the marginal conditional distributions are modeled first; then copula theory is applied to approximate the dependence among the residual terms of the marginals.

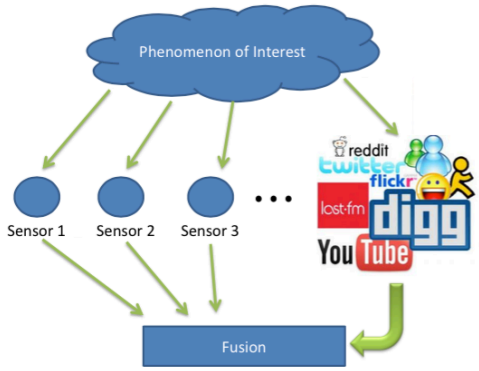

Fig. 3.1: The network of traditional sensors and social media.

3.2 Problem Formulation

We consider a network consisting of traditional sensors and a set of “human sensors" providing a massive amount of accessible social media data, as shown in Figure 3.1. Traditional sensors take successive measurements about the phenomenon of interest over a time interval $l = 1, \ldots, L$, while social media data about the same event is successively collected over the same time interval. Informative features, such as volume and sentiment, are, then, extracted from the social media data for inference purposes.

We assume that there are a total of N data sources in the network which includes sensors and extracted features from social media. We use $x_{n,l}, n \in \{1, \ldots, N\}$ to denote the observation of traditional sensor n or the feature n of the social media data, at time instant l. The inference task is conducted at the fusion center based on the N-dimensional time series $\{x_{1,l}, \ldots, x_{N,l}\}_{l=1}^{L}$ obtained from the sensors and social media. A statistical model is needed to characterize the vector stochastic process before any inference task, such as detection, estimation, and prediction, can be conducted. Since sensor data and social media data provide noisy observations of the same phenomenon, they are dependent across modalities. As discussed in Section 3.1, observations from different sources, especially when including both sensors and social media, are very likely to be heterogeneous where the heterogeneity under the assumption

that the N-dimensional random vector are i.i.d. in time is defined by Iyengar et al. in [51] as follows

Definition 3.1 (S. Iyengar [51] Def 1)**.** *A random vector $\{X_1, \ldots, X_N\}$ governing the joint statistics of N-variate data set is termed as heterogeneous if the marginals X_n are non-identically distributed.*

Considering the temporal dependence, which exists in most consecutive measurements but neglected for simplification, we extend the above definition of heterogeneity to a vector stochastic process as follows

Definition 3.2. *A vector stochastic process $\{X_{1,l}, \ldots, X_{N,l}\}_{l=1}^{L}$ with dimension N is heterogeneous if the conditional marginal distributions $f_n(x_{n,l}|\, \mathbf{x}_n^{l-1})$ are non-identical, where $\mathbf{x}_n^{l-1} := \{x_{n,l}\}_{l=1}^{l-1}$.*

In all, there are two main properties of social data and sensor data that need to be captured by the statistical model:

- Two-dimensional dependence: temporal and intermodal dependence.
- Heterogeneity: disparate conditional marginal distributions.

To accommodate the above requirement of the statistical characterization of sensor data and social media data, in this chapter we formulate a conditional copula-based approach to approximate the joint distribution. The following assumption is made about the observations.

Assumption 3.1. *The temporal dynamics of each time series does not depend on any other time series, i.e.,*

$$f_n(x_{n,l}|\mathcal{I}^{l-1}) = f_n(x_{n,l}|\, \mathbf{x}_n^{l-1}), \forall n = 1, \ldots, N$$

where $\mathcal{I}^{l-1} := \{x_{1,l}, \ldots, x_{N,l}\}_{l=1}^{l-1}$ denotes the entire observation set before time instant l.

3.3 Copula-based Multivariate Dynamic Models

Before modeling the dependence structure between the N time series using copulas, we need to model their conditional marginal distributions.

3.3.1 Conditional Marginal Distributions

We focus on the first two moments of the marginal conditional distributions. Assuming that the first two moments exist, we apply the following structure, which is commonly used to represent time series with time-varying conditional mean and time-varying conditional variance [82], to model each marginal

$$x_{n,l} = \mu_n(\mathbf{x}_n^{l-1}\,|\boldsymbol{\theta}_n) + \sigma_n(\mathbf{x}_n^{l-1}\,|\boldsymbol{\theta}_n)\epsilon_{n,l}, \quad n = 1,\ldots,N; l = 1,\ldots,L \tag{3.1}$$

where $\boldsymbol{\theta}_n$ represents the set of unknown parameters and $\epsilon_{n,1}, \epsilon_{n,2}, \ldots, \epsilon_{n,L}$ is a sequence of i.i.d. random variables with zero mean and unit variance, but without a specified PDF.

The model in (3.1) allows each time series to have a time-varying conditional mean and time-varying conditional variance.

The Auto Regressive Moving Average ARMA model [72] provides a description of the conditional mean part of Eq. (3.1). An ARMA(P, Q) model, where P is the order of the autoregressive part and Q is the order of the moving average part, is written as

$$x_{n,l} = \alpha_0 + \sum_{i=1}^{P} \alpha_i x_{n,l-i} + \sum_{j=1}^{Q} \beta_j z_{n,l-j} + z_{n,l} \quad n = 1,\ldots,N; l = 1,\ldots,L \tag{3.2}$$

where $z_{n,l}$ are the error terms that are generally assumed to be i.i.d random variables and with zero mean.

The Generalized autoregressive conditional heteroskedasticity (GARCH) model [16, 38] characterizes the conditional variance of the error term $z_{n,l} = \sigma_{n,l}\epsilon_{n,l}$ by imposing alternative

Table 3.1: Behaviors of theoretical ACF and PACF

Model	ACF	PACF
AR(P)	exponential decay and/or damped sinusoid	cuts off after lag P
MA(Q)	cuts off after lag Q	exponential decay and/or damped sinusoid
ARMA(P,Q)	exponential decay and/or damped sinusoid	exponential decay and/or damped sinusoid

parameters to capture serial dependence on the past sequence of observations as

$$\sigma_{n,l}^2 = a_0 + \sum_{i=1}^{M} a_i \sigma_{n,l-i}^2 + \sum_{j=1}^{N} b_j z_{n,l-j}^2, \quad n = 1, \ldots, N; l = 1, \ldots, L \tag{3.3}$$

In order to roughly determine the orders of both AR and MA parts of the model, we consider the autocorrelation function (ACF) and partial autocorrelation function (PACF).[1] The theoretical behaviors of the ACF and PACF are summarized in Table 3.1 [72]. As can be seen in Eq. (3.3), if an ARMA model is assumed for the error variance $\sigma_{n,l}^2$, the model is a GARCH model. Thus, the approximate order of GARCH(M, N) can be obtained by checking the ACF and PACF of the squared residuals $z_{n,l}^2$.

As to determining the optimal orders of the ARMA and GARCH models, there are a variety of model selection criteria such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC), which are measures of the relative quality of a statistical model for a given set of data. That is, given a collection of models for the data, AIC/BIC estimate the quality of each model, relative to the other models. They are formally defined as

$$\begin{aligned} AIC &= -2 \log \mathcal{L} + 2k \\ BIC &= -2 \log \mathcal{L} + k \ln L \end{aligned} \tag{3.4}$$

where $\mathcal{L}$ is the maximized value of the likelihood function for the model; k is the number of

[1] PACF is the autocorrelation after adjusting for a common factor.

parameters in the model (i.e. k is the number of degrees of freedom) and L is the number of observations, or equivalently, the sample size.

After fitting each time series to the selected models, the parameters of the ARMA-GARCH models are obtained and the estimated standardized residuals are

$$\hat{\epsilon}_{n,l} = \frac{x_{n,l} - \mu_n(\mathbf{x}_n^{l-1} | \hat{\theta}_n)}{\sigma_n(\mathbf{x}_n^{l-1} | \hat{\theta}_n)} \tag{3.5}$$

where $\hat{\theta}_n$ is the estimate of the parameters of each time series, for all $n = 1, \ldots, N$.

While a scalar GARCH model is used to capture the time-varying variance of each individual time series in this chapter, a parametric copula is used to model the intermodal dependence between different time series.

3.3.2 Estimation and Inference for Copula Models

In this subsection, we introduce the inference for copula-based multivariate models. The conditional copula is the joint distribution of the probability integral transforms of the standardized residuals. We consider the parametric copula model and its parameter to be invariant with time.

An important benefit of using copulas to construct multivariate models is that the models used in the marginal distributions need not be of the same type as the model used for the copula. One exciting possibility that this allows is semi-parametric estimation of the marginal distributions, combined with parametric estimation of the copula. Such a model avoids the curse of dimensionality by only estimating the one-dimensional marginal distributions non-parametrically, and then estimating the copula parametrically. In the semi-parametric model, the marginal distributions are modeled non-parametrically, using distributions such as the empirical distribution function (EDF) and a parametric model is used for intermodal dependence. When copula c is used to characterize the dependence structure, its corresponding parameters

are estimated using the MLE based approach, as follows

$$\hat{\boldsymbol{\phi}} = \arg\max_{\boldsymbol{\phi}} \sum_{l=1}^{L} \log c(\hat{u}_{1,l}, \ldots, \hat{u}_{N,l}|\boldsymbol{\phi}) \tag{3.6}$$

where

$$\hat{u}_{n,l} = \hat{F}_{\epsilon_n}(\epsilon_{n,l}), \quad \forall n = 1, \ldots, N \tag{3.7}$$

We consider the nonparametric estimate of the CDF F_n using the EDF [2]:

$$\hat{F}_{\epsilon_n}(\epsilon) = \frac{1}{L+1} \sum_{l=1}^{L} \mathbb{1}_{\{\hat{\epsilon}_{n,l} \leq \epsilon\}} \tag{3.8}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function and the estimated standardized residual $\hat{\epsilon}_{n,l}$ is given in Eq. (3.5).

When the true dependence structure is unknown, the best copula model needs to be selected from a finite set of predefined copula density functions $\mathcal{C} = \{c_1, \ldots, c_m\}$. There are several copula selection approach that can be applied, such as AIC and BIC based approaches. We use an MLE based approach, since for fixed N, different copulas do not have much difference in parameter dimensions. Thus, the best copula c^* is selected as follows

$$c^* = \arg\max_{c_m \in \mathcal{C}} \sum_{l=1}^{L} \log c_m(\hat{u}_{1,l}, \ldots, \hat{u}_{N,l}|\hat{\boldsymbol{\phi}}) \tag{3.9}$$

where $\hat{\boldsymbol{\phi}}$ is obtained using MLE in Eq. (3.7).

We apply the copula-based characterization of multivariate time series to social media data

[2]The definition of EDF is scaled by $1/(L+1)$ rather than $1/L$. This has no effect asymptotically, and it is useful in keeping the estimated CDF away from the boundaries of the unit interval, where some copula models diverge.

assisted stock prediction as an application of our methodology in the next section.

3.4 Stock Prediction with Google Trends Data

The analysis and forecasting of stock market behavior has been a focus of academics and practitioners alike. A model that accounts for investor attention can provide a better explanation of the stock behavior and can also contain useful information to forecast stock market. Google Trends (GT) serves as a good “attention indicator" that is measured through social media. GT is a public web facility of Google Inc., based on Google Search, that shows how often a particular search-term is entered. When one searches for a term on GT, one sees a graph showing its popularity over time. The numbers reflect how many searches have been done for a particular term. [3]

It is reasonable to assume that the GT volume, indicating the level of people’s interest, is correlated with stock price. If such correlation exists, stock price could possibly be better predicted with the assistance of GT data. The main focus of this section is to predict the stock price with the assistance of GT data, taking Apple Inc. as an example. We collect the weekly stock price data, which serves as a surrogate for sensor data in our example, and weekly GT data [4] using the search keywords “Apple Inc." + “Apple stock" [5] from Oct. 3 2004 to Jun. 6 2014. The normalized stock price and GT data can be seen in Figure 3.2.

Most financial studies involve returns, instead of prices, of assets mainly because return series has more attractive statistical properties [109]. Thus, we convert the stock price data and GT data to return series [6], as shown in Figure 3.3, for the following analysis to be conducted.

Let the return series of stock be denoted as $\{x_{1,l}\}_{l=1}^{L}$ and the return series of GT be denoted

[3]The numbers don’t represent absolute search volume numbers, because the data is normalized and presented on a scale from 0-100. Each point on the graph is divided by the highest point and multiplied by 100.

[4]Data Source: Google Trends (www.google.com/trends).

[5]“ + ” represents that the results can include searches containing the words Apple Inc. OR Apple stock.

[6]Let $\{y_l\}_{l=1}^{L}$ denote the original time series, the returns in this chapter are calculated as follows: $\nabla y_l = y_l - y_{l-1}$

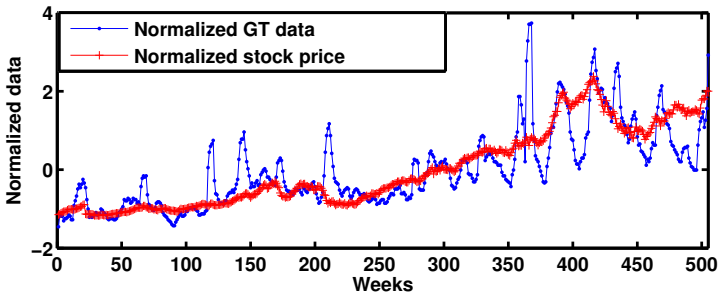

Fig. 3.2: Normalized weekly stock price and google search volume of Apple Inc. (from Oct. 3 2004 to Jun. 6 2014).

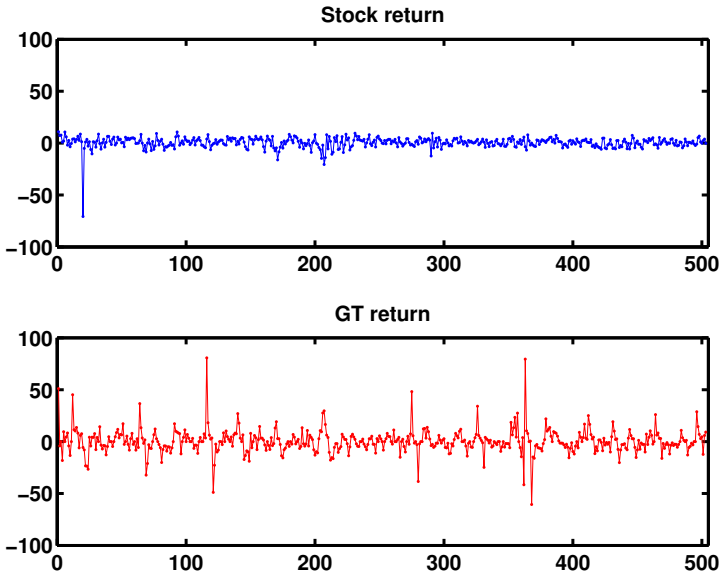

Fig. 3.3: Stock returns and GT returns (multiplied by 100).

as $\{x_{2,l}\}_{l=1}^{L}$, and in this example $L = 505$. We aim at predicting the stock returns based on its historical data with the assistance of the GT returns i.e., $\hat{x}_{1,l+1} = g(\mathbf{x}_1^{(l)}, \mathbf{x}_2^{(l)})$, by first constructing the conditional joint distribution of $f(x_{1,l+1}, x_{2,l} | \mathbf{x}_1^{(l)}, \mathbf{x}_2^{(l-1)})$. It has to be noted that the two time series are intentionally misaligned by one time instant, and it will be justified later by the non-negligible dependence among the standardized residuals of the misaligned returns.

We first model each of the two returns individually. By observing the patterns of ACF and

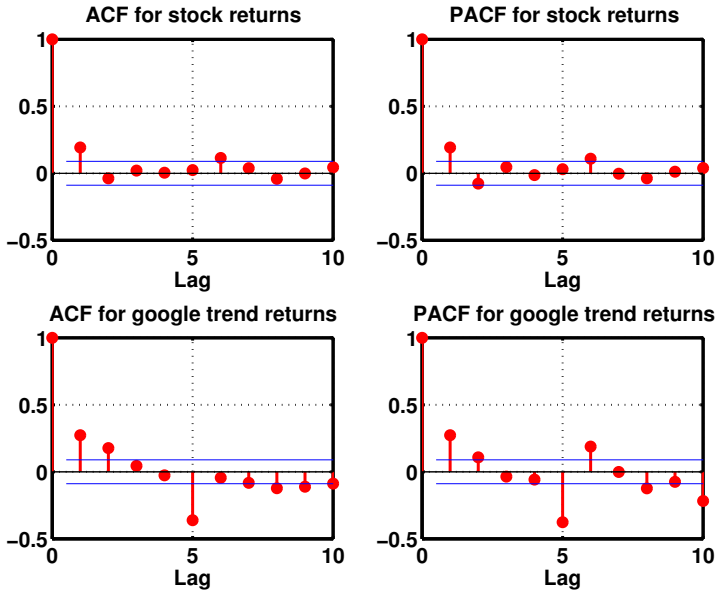

Fig. 3.4: ACF and PACF of stock returns and GT returns

PACF in Figure 3.4, we know that, according to Table 3.1, both of the two returns have non-zero AR lag and MA lag, and an ARMA up to the order $(5, 5)$ is good enough for modeling the conditional mean part in Eq. (3.1). Using AIC, an ARMA(5,5) model is selected for stock returns and ARMA(4,5) is selected for GT returns. When BIC is applied, ARMA(5,5) model is chosen for the stock returns and ARMA(1,5) model is selected for GT returns.

We conduct Engle's ARCH test [7] for the existence of time-varying conditional variance on the residual series $\{z_{n,1}, \ldots, z_{n,L}\}$. For the stock returns, the null hypothesis is accepted indicating that the variance of the residual series is not time-varying. The result is also supported by Figure 3.5, in which the stock returns do not have either GARCH term (M) or ARCH term (N), namely constant variance model can be applied to the residuals of stock returns. As to the GT returns, GARCH models up to order $(5, 5)$ are considered, and by AIC, a GARCH(1,1) is selected, while a GARCH(0,1) is selected by BIC.

After the marginal models are selected and corresponding parameters are estimated, the estimated standardized residuals $\{\hat{\epsilon}_{1,1}, \ldots, \hat{\epsilon}_{1,L}\}$ and $\{\hat{\epsilon}_{2,1}, \ldots, \hat{\epsilon}_{2,L}\}$ are obtained through Eq.

[7] Engle's ARCH test is a Lagrange multiplier test to assess the significance of ARCH effects [38].

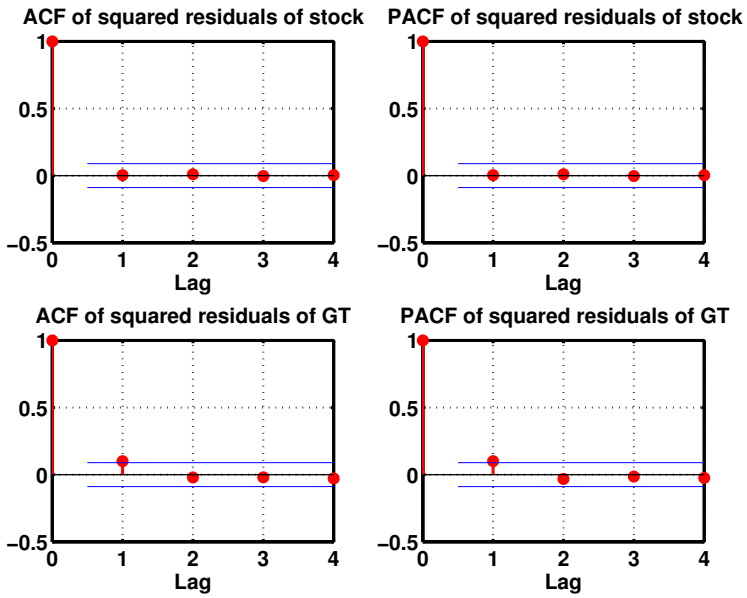

Fig. 3.5: ACF and PACF of the squared residuals of the stock returns and GT returns

(3.5). By checking the Kendall's τ [8] for different alignments of the two data sets, i.e., $\{\hat{\epsilon}_{1,1+d}, \ldots, \hat{\epsilon}_{1,L}\}$ and $\{\hat{\epsilon}_{2,1}, \ldots, \hat{\epsilon}_{2,L-d}\}$, we find that $d = 1$ renders the largest $\tau = 0.047$. Thus, the strongest dependence exists between $\epsilon_{1,l+1}$ and $\epsilon_{2,l}$. The dependence between the standardized residuals of stock data at $l+1$ and GT data at l allows us to utilize the current GT data to predict the stock returns of the next week. And the best copula that is selected to model the dependence between the standardized residuals is Frank copula.

It is known that the optimal estimator that minimizes the mean square error (MSE) is the conditional expectation as follows

$$\begin{aligned} \hat{x}_{1,l+1} &= \mathbb{E}[X_{1,l+1} | \, \mathbf{x}_1^l, \mathbf{x}_2^l] \\ &= \mu_1(\mathbf{x}_1^l \, | \hat{\theta}_1) \\ &+ \sigma_1 \int f_{\epsilon_1}(\epsilon_1) c^*(F_{\epsilon_1}(\epsilon_1), F_{\epsilon_2}(\epsilon_{2,l}) | \hat{\boldsymbol{\theta}}_d) d\epsilon_1 \end{aligned} \tag{3.10}$$

where σ_1 is not time-variant since we have shown that there is no conditional heteroskedasticity

[8] As defined earlier, Kendall's τ is a non-parametric rank-based measure of dependence ranging from -1 to 1.

Table 3.2: MSE of different prediction approaches

Approach	AIC & Frank Copula	BIC & Frank Copula	$c* = 1$
MSE	6.1800	6.2348	6.5960

in the stock returns.

In each of the 30 trials that have been conducted, we use the first l data points, $\{x_{1,2}, \ldots, x_{1,l}\}$ and $\{x_{2,1}, \ldots, x_{2,l-1}\}$ as the training set for the estimation of the model parameters $\{\boldsymbol{\theta}, \boldsymbol{\phi}\}$, and use the next data point $x_{1,l+1}, x_{2,l}$ as testing data to check the square error of our estimator. The MSE is calculated as follows

$$MSE = \frac{1}{30} \sum_{l=476}^{505} (x_{1,l} - \hat{x}_{1,l})^2 \tag{3.11}$$

Compared with the approach that assumes intermodal independence between the two returns, i.e., setting $c^* = 1$ in Eq. (3.10), we are able to reduce the MSE by around 5% as shown in Table 3.2 by capturing the intermodal dependence and exploiting the dependence in the process of prediction.

3.5 Flu Prediction with Google Trends Data

In this section, we apply our proposed copula-based characterization for heterogeneous time series to flu prediction using GT data and Influenza-Like Illness (ILI) data. The weekly GT data is obtained from Google Trends using flu-related keywords: “Flu" + “Flu symptoms". The weekly ILI data is obtained from the website of Centers for Disease Control and Prevention (CDC) [9] and it represents the total number of ILI reported by U.S. World Health Organization (WHO) and the National Respiratory and Enteric Virus Surveillance System (NREVSS) collaborating Laboratories. Both GT data and ILI data are collected from Jan. 4 2004 to Feb. 8 2015. In this case, ILI data serves as the “sensor" data. We are interested in predicting the

[9] Data source: http://www.cdc.gov/

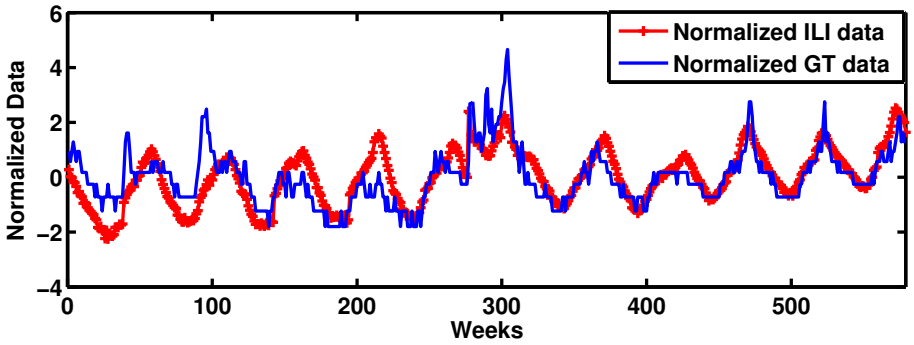

Fig. 3.6: Normalized weekly Influenza-Like Illness (ILI) data and Google Trends (GT) data (from Jan. 4 2004 to Feb. 8 2015).

ILI data using historical ILI data and GT data. The intuition is that when the population affected by flu increases, flu-related search volume increases with it, and so does the number of ILI reported by U.S. WHO and NREVSS. GT data and ILI data are observations of the same phenomenon, thus correlated. The normalized GT data and ILI data are shown in Figure 3.6 and the size of the raw data set is $L^{'} = 579$.

In the previous example of stock prediction, the serial dependence of the current observation on the previous observations is strongest for the immediate past and followed by a decaying pattern as we move further back in time. However, for the flu data, including ILI data and GT data, as shown in Figure 3.6, the dependence shows a repeating, cyclic behavior. The flu data of one year (52 weeks) is somewhat similar to that of the previous year. In other words, the flu data in one week shows a strong correlation not only with the data in last several weeks but also with the data in the same week of the previous years. This cyclic pattern, or as more commonly called *seasonal pattern*, can be effectively used to further improve the forecasting performance. Thus, we will extend the use of ARMA models to seasonal data for characterizing each marginal distribution in this section.

The most important structural issue to recognize about seasonal time series is that if the season is s periods long, then observations that are s time intervals apart are alike. In our case, $s = 52$ weeks as shown in Figure 3.6. We therefore have two relationships going on

simultaneously (1) between observations for successive weeks within the same year, and (2) between observations for the same week in successive years. Therefore, we need to build two time series models for both of them and then combine the two.

The model building process for seasonal data is similar to that used for nonseasonal data. First, if the data is not stationary, we need to preprocess it to make the data stationary so that the autocorrelation dies out quickly. For seasonal data, we may not only use the regular difference $\nabla y_t = y_t - y_{t-1}$ but also a seasonal difference $\nabla_s y_t = y_t - y_{t-s}$. And in our case, it is shown that we need both to obtain an ACF that dies out sufficiently quickly. For ILI data and GT data, with an $s = 52$ seasonality, we use the combined $\nabla\nabla_{52} y_t = \nabla(y_t - y_{t-52}) = y_t - y_{t-1} - y_{t-52} + y_{t-53}$ to obtain the combined difference time series $\{x_{1,l}\}_{l=1}^{L}$ and $\{x_{2,l}\}_{l=1}^{L}$, respectively, where $L = 526$. For simplicity of presentation, we will refer to the two combined difference time series as ILI data and GT data, in the remainder of this section.

Having transformed the data to a stationary form using the $\nabla\nabla_{52} y_t$ difference operation, we are ready to identify the seasonal time series. The methodology for identifying stationary seasonal models is a modification of the one used for regular ARMA as described in Section 3.4. The main patterns in ACF and PACF that we are looking for in nonseasonal ARMA models have been summarized in Table 3.1. And we will use the same table for seasonal ARMA models as well. The difference is that we are looking for signs of two models: one seasonal and one regular week-to-week model.

For the ACF of ILI data, it can be seen from Figure 3.7 that there is an isolated significant negative spike at lag 52. It is a sign of the MA model applied to the 52-week seasonal pattern. For the regular model, it can be seen from Figure 3.7 that the autocorrelation between consecutive weeks seems insignificant. The negative spikes at lag 52 and 104 suggest the need for a first and second order autoregressive in the seasonal model. Since the regular ARMA model is not necessary for the ILI data as shown in 3.7, we choose to use only the seasonal one, i.e., ARMA$(2, 1)_{52}$. Similarly, an ARMA$(1, 1)_{52}$ is applied to model the GT data. As to the het-

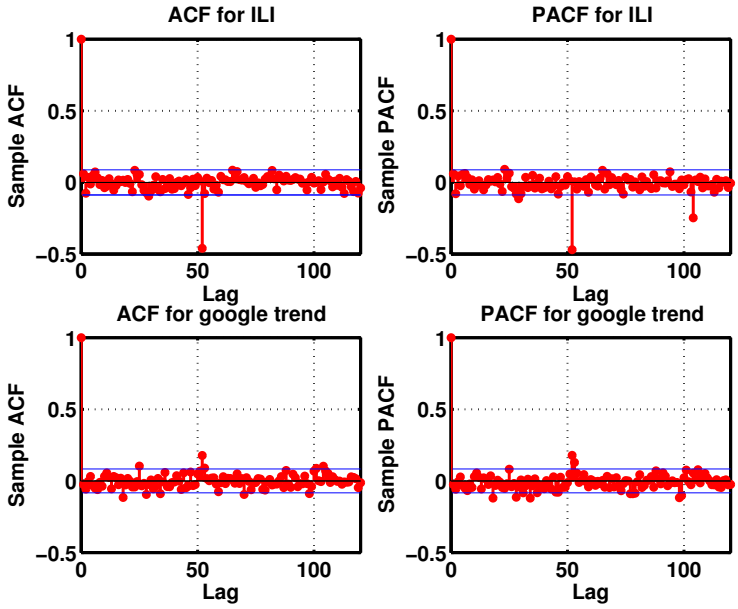

Fig. 3.7: ACF and PACF of ILI data and GT data (after taking the combined difference)

eroskedasticity, it can be seen from Figure 3.8 that there is no significant heteroskedasticity for both ILI data and GT data. Thus, we choose to assume constant variance for the residual terms of the two time series.

After the marginal models are selected and corresponding parameters are estimated, the estimated standardized residuals $\{\hat{\epsilon}_{1,1}, \ldots, \hat{\epsilon}_{1,L}\}$ and $\{\hat{\epsilon}_{2,1}, \ldots, \hat{\epsilon}_{2,L}\}$ are obtained through Eq. (3.5). The alignment of the two data sets $\{\hat{\epsilon}_{1,1+d}, \ldots, \hat{\epsilon}_{1,L}\}$ and $\{\hat{\epsilon}_{2,1}, \ldots, \hat{\epsilon}_{2,L-d}\}$ with $d = 1$ renders the $\tau = 0.0215$. The dependence between the standardized residuals of ILI data at $l+1$ and GT data at l allows us to utilize the current GT data to predict the ILI data of the next week. And the best copula that is selected to model the dependence between the standardized residuals is Student's t copula with the degree of freedom $\nu = 4$.

The optimal estimator that minimizes the mean square error (MSE) is the conditional expectation as shown in (3.10). In each of the 100 trials, we use the first l data points, $\{x_{1,2}, \ldots, x_{1,l}\}$ and $\{x_{2,1}, \ldots, x_{2,l-1}\}$ as the training set for the estimation of the model parameters $\{\boldsymbol{\theta}, \boldsymbol{\phi}\}$, and use the next data point $x_{1,l+1}, x_{2,l}$ as testing data to check the square error of our estimator. Compared with the approach that assumes intermodal independence between the two returns,

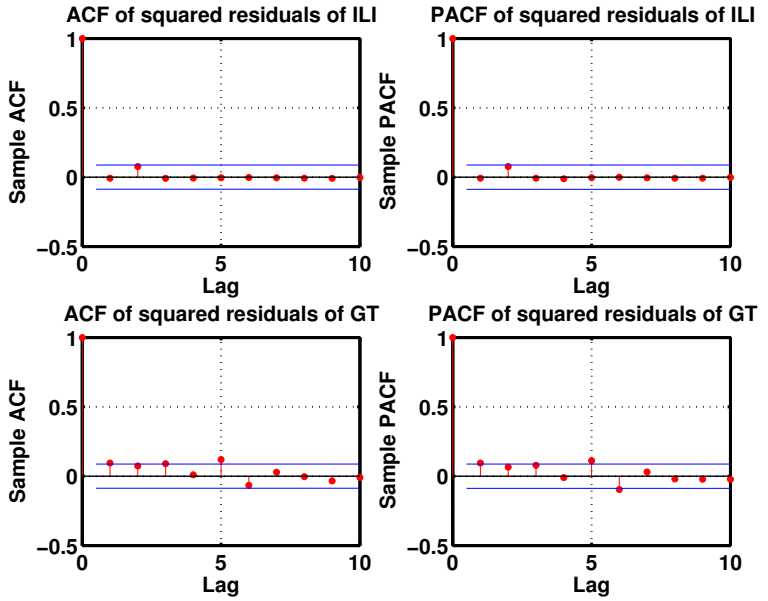

Fig. 3.8: ACF and PACF of the squared residuals of ILI data and GT data (after taking the combined difference)

Table 3.3: MSE of different prediction approaches

Approach	t Copula	$c* = 1$
MSE	8.960	9.090

i.e., setting $c^* = 1$ in Eq. (3.10), we are able to reduce the MSE by 1.43% as shown in Table 3.3 by capturing the intermodal dependence and exploiting the dependence in the process of prediction. Even though the reduction of MSE by 1.43% dose not seem to be very significant, considering the fact that actual values of the ILI data are around ten thousand, the result is important. For the prediction method which assumes independence between ILI data and GT data, the square root of MSE corresponding to the raw data is 1504, which can be roughly interpreted as that the predicted ILI data deviate from the true value by 1504 on an average. By our approach, the square root of MSE corresponding to the raw data is 1494. It means that the predicted numbers of Influenza-Like Illness using our method are closer to the true value by $1504 - 1494 = 10$ people on an average, when compared to the method under independence assumption.

3.6 Summary

In the networks of sensors and social media, temporal dependence may exist, along with intermodal dependence. We proposed a copula-based model for multivariate time series which is very flexible in capturing a wide range of temporal and spatial dependence structures. It provides a powerful tool for characterizing heterogeneous time series (sensor data and social data) by accommodating very complex and nonlinear dependence structures. By applying such a model to jointly characterize the GT returns and stock returns of Apple Inc., we are able to have a better prediction of the stock. Its application in other social media data assisted inference tasks, such as flu prediction, was also investigated.

CHAPTER 4
DISTRIBUTED CLASSIFICATION UNDER DEPENDENT OBSERVATIONS

4.1 Motivation

Classification using multiple sensors is generally more reliable in that it yields more accurate results, and has been widely studied in several engineering applications like, target recognition, and identification. Distributed processing approaches for classification are desired in wireless sensor networks (WSNs) because gathering all sensors' observations at the FC requires large communication bandwidth. In this approach, the sensors process and analyze their observed raw data and transmit only the compressed information to the FC, which then generates the final decision. Distributed hypothesis testing schemes for classification have received significant attention, but most research has focused on cases where the observations at different sensors are independent [30, 45, 117]. However, when sensors observe the same random signals, their observations may not be independent.

The Automatic Modulation Classification (AMC) problem serves as a good example of the above scenario, where under each modulation scheme (hypothesis), the communication sig-

nal can be viewed as a discrete random variable taking values from the corresponding set of constellation symbols. AMC is a signal processing method that is used to classify the modulation scheme corresponding to the received noisy communication signals and plays a key role in various civilian and military applications. Extensive research has been done on AMC methods with a single receiver [31, 32, 46, 81, 92], whose performance depends heavily on the channel quality. Thus, network centric methods for AMC using multiple sensors have been motivated and investigated [40, 79, 80, 95, 124, 129]. In many cases, due to the scarcity of transmission resources including channel bandwidth and local sensors' energy, distributed processing of locally sensed signals is desirable. Accordingly, local decision rules and the fusion rule in such environments are investigated in [95, 124, 129]. In [124], each sensor conducts a test based on the likelihood ratio of its observations, which, according to [48], is optimal only with conditionally independent sensor observations. To the best of our knowledge, no work has tackled the problem of distributed modulation classification that considers conditionally dependent observations, which is the topic studied in this chapter.

Some recent efforts on distributed detection with conditionally dependent observations are discussed in [23, 50, 91, 131], where [50] focuses on only the fusion aspect of the problem and [23] emphasizes a very general theoretical framework. In this chapter, we investigate the optimal rules at the sensors and the FC in the Bayesian framework with dependent observations for the distributed classification problem. Our approach is based on the introduction of a new "hidden" random variable as proposed in [23], through which a hierarchical conditional independence model is built. We derive the necessary condition for optimal decision rules at the sensors and the FC, based on which, an iterative optimization algorithm is proposed. We also address the implementation issue of the iterative algorithm by discretizing the observation space of the local sensors.

4.2 Problem Formulation

We consider a distributed hypothesis testing problem in wireless sensor networks. Suppose there are K candidate hypotheses, represented by $H_0, \ldots, H_{K-1}$, with prior probabilities $\pi_0, \ldots, \pi_{K-1}$. In the example of AMC, the candidate hypotheses correspond to different modulation schemes, as will be demonstrated in a later section of this chapter. The signal is represented by a discrete random variable s. Under H_i, s takes value from the symbol set $\mathcal{S}_i$, i.e., $s \in \mathcal{S}_i = \{I_i^1, \ldots, I_i^{M_i}\}$, where $I_i^{m_i}$ is the m_i-th symbol. It is assumed that the probability of each symbol in $\mathcal{S}_i$ is equal under given hypothesis H_i. Thus, the probability mass function (PMF) of signal s under each hypothesis is given as follows

$$H_i : P(s = I_i^{m_i}) = \frac{1}{M_i} \quad \forall I_i^{m_i} \in \mathcal{S}_i \tag{4.1}$$

for $i = 0, \ldots, K-1$. It is noted that the symbol sets under different hypotheses may overlap, i.e., $\mathcal{S}_i \cap \mathcal{S}_j \neq \emptyset, \forall i \neq j$.

We consider a general signal reception scenario with N sensors where each sensor takes L observations for decision making. The received observation of sensor $n, n = 1, \ldots, N$ at time $l, l = 1, \ldots, L$ can be written as:

$$x_{nl} = h_n s_l + w_{nl} \tag{4.2}$$

where h_n is the channel gain and $\{s_l\}_{l=1}^{L}$ is the discrete random signal sequence. We assume w_{nl} to be independent and identically distributed (i.i.d) noise whose distribution is known. Thus, the PDF of x_{nl} conditioned on the hypothesis H_i and the symbol s_l can be obtained and written as $f_n(x_{nl}|s_l, H_i)$, where $f_n(\cdot|H_i)$ denotes the PDF under hypothesis H_i.

As shown in Figure 4.1, each sensor n locally processes the observation sequence $\mathbf{x}_n := [x_{n1}, \ldots, x_{nL}]^T$ and makes a decision $u_n = \gamma_n(\mathbf{x}_n)$, where $u_n \in \mathcal{K} := \{0, \ldots, K-1\}$, where $\gamma_n(\cdot)$ is the decision rule at sensor n. When the FC receives the local decisions $\mathbf{u} :=$

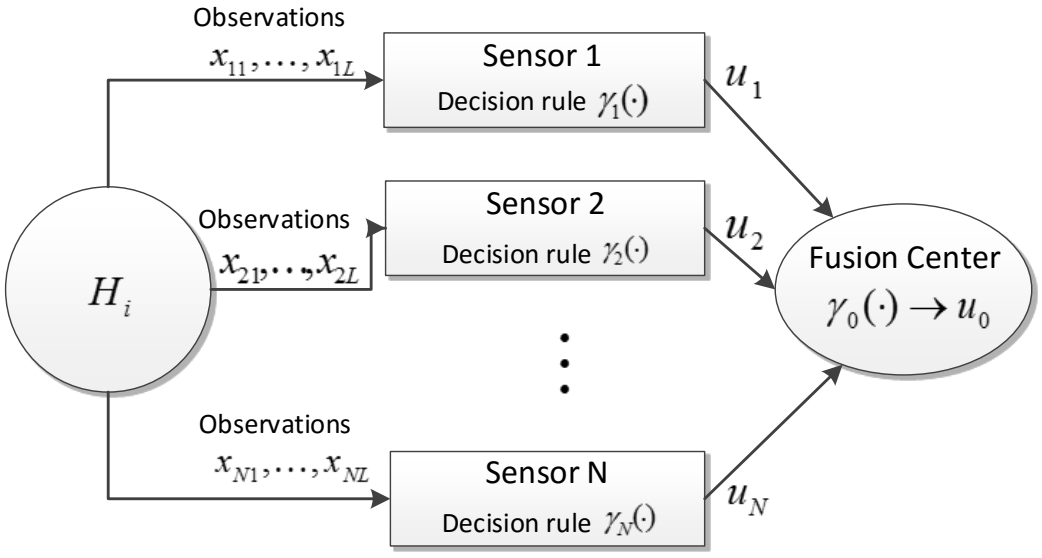

Fig. 4.1: Distributed classification using a wireless sensor network

$[u_1, \ldots, u_N]$ from all the sensors, it makes the final decision $u_0 = \gamma_0(\mathbf{u})$, where $u_0 \in \mathcal{K}$ and $\gamma_0(\cdot)$ is the fusion rule. We assume that the true underlying hypothesis remains unchanged during the collection of L observations.

Our goal is to design the set of local decision rules and the fusion rule, i.e., $\boldsymbol{\gamma} := \{\gamma_1 \ldots, \gamma_N, \gamma_0\}$ to maximize the classification performance, in terms of probability of correct classification P_c, considering the fact that the sensors observing the same discrete random signals have dependent observations, i.e.,

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_N | H_i) \neq \prod_{n=1}^{N} f(\mathbf{x}_n | H_i) \tag{4.3}$$

4.3 Design of Optimal Sensor Rules and Fusion Rule

In the distributed classification system, the random variables involved form the following Markov chain:

$$H \rightarrow \mathbf{X} \rightarrow \mathbf{U} \rightarrow U_0 \tag{4.4}$$

where H represents the hypotheses, $\mathbf{X}$ is the observation matrix, $\mathbf{U}$ is the vector of sensor decisions, and U_0 is the final decision. Independence among $\mathbf{X}$ conditioned on H makes the derivation of decision rules easy by allowing factorization of $f(\mathbf{X}|H)$, but it does not hold here as discussed in the previous section. However, according to the observation model in (4.2), for a given symbol s_l which takes a value from the set $\mathcal{S} := \cup_{i=0}^{K-1}\mathcal{S}_i$, the variables $[X_{1l}, \ldots, X_{Nl}]$ are independently distributed, for all $l = 1, \ldots, L$. Thus, we define a random vector $\mathbf{Y} := [s_1, \ldots, s_L]$ to represent the underlying symbols. And by introducing $\mathbf{Y}$ between H and $\mathbf{X}$, into the Markov chain in (4.4), the following results hold:

1. Given $\mathbf{Y}$, $\mathbf{X}$ is independent of H, namely, the following Markov chain is valid:

$$H \rightarrow \mathbf{Y} \rightarrow \mathbf{X} \rightarrow \mathbf{U} \rightarrow \mathbf{U}_0 \tag{4.5}$$

2. The introduction of random variable $\mathbf{Y}$ makes sensors' observations conditionally independent of each other, and thus facilitates the design of sensor decision rules. In other words, $\mathbf{X}_1, \ldots, \mathbf{X}_N$ are independent conditioned on $\mathbf{Y}$, i.e.,

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_N|\mathbf{Y}) = \prod_{n=1}^{N} f(\mathbf{x}_n|\mathbf{Y}). \tag{4.6}$$

We aim to derive the sensor decision rules and the fusion rule that maximize the probability of correct classification P_c

$$P_c = \sum_{u_0=0}^{K-1} \sum_{i=0}^{K-1} c_{u_0,i} P(u_0|H_i)\pi_i \tag{4.7}$$

where, $c_{u_0,i}$ indicates the cost of deciding u_0 when the true hypothesis is H_i. We consider the particular cost assignment where $c_{u_0,i} = 1$ if and only if $u_0 = i$, otherwise, $c_{u_0,i} = 0$, which means that a correct classification occurs when the decision at the FC matches with the true hypothesis. Thus, P_c can be further simplified as

$$P_c = \sum_{u_0=0}^{K-1} P(u_0|H_{u_0})\pi_{u_0} \tag{4.8}$$

We can express P_c with respect to sensor n as follows:

$$\begin{aligned} P_c &= \int_{\mathbf{X}} \sum_{\mathbf{u}} \sum_{u_0=0}^{K-1} \pi_{u_0} P(u_0|\mathbf{u}) P(\mathbf{u}|\mathbf{r}) f_{u_0}(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbf{X}_n} \sum_{u_n} P(u_n|\mathbf{x}_n) g_n(u_n, \mathbf{x}_n) d\mathbf{x}_n \end{aligned} \tag{4.9}$$

where $\mathbf{x} := [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ and

$$g_n(u_n, \mathbf{x}_n) = \sum_{\mathbf{u}^n} \sum_{u_0=0}^{K-1} \pi_{u_0} P(u_0|\mathbf{u}^n, u_n) \int_{\mathbf{R}^n} P(\mathbf{u}^n|\mathbf{x}^n) f(\mathbf{x}|H_{u_0}) d\mathbf{x}^n \tag{4.10}$$

where a vector (matrix) with a superscript denotes it without the nth element (column), for example, $\mathbf{X}^n$ represents the vector $\mathbf{X} \setminus X_n = [X_1, \ldots, X_{n-1}, X_{n+1}, \ldots, X_N]$.

To maximize P_c, the optimal decision rule at sensor n given fixed decision rules at all the other sensors and the FC is to make a decision u_n such that $g_n(u_n, \mathbf{x}_n)$ is maximized, namely

$$\gamma_n(\mathbf{x}_n) = \arg\max_{u_n} g_n(u_n, \mathbf{x}_n) \tag{4.11}$$

for all $\mathbf{x}_n$. This is because for given $\mathbf{x}_n$, $P(u_n|\mathbf{x}_n) = 1$ only for $u_n = \gamma_n(\mathbf{x}_n)$, otherwise $P(u_n|\mathbf{x}_n) = 0$, thus the decision rule in (4.11) maximizes the probability of making a correct decision when $\mathbf{x}_n$ is observed.

With the introduction of the hidden random vector $\mathbf{Y}$, the joint distribution of all sensors' observations $\mathbf{x}$ conditioned on hypothesis H_i, i.e., $f(\mathbf{r}|H_i)$ ($f(\mathbf{x}|H_{u_0})$ in (4.10)), can be written as follows:

$$\begin{aligned} f(\mathbf{x}|H_i) &= \sum_{\mathbf{Y}} P(\mathbf{Y} = \mathbf{y}|H_i) \prod_{n=1}^{N} f(\mathbf{x}_n|\mathbf{Y} = \mathbf{y}) \\ &= \sum_{\mathbf{Y}} \left(\frac{1}{M_i}\right)^L \prod_{n=1}^{N} f(\mathbf{x}_n|\mathbf{Y} = \mathbf{y}) \end{aligned} \tag{4.12}$$

where $\sum_{\mathbf{Y}} = \sum_{\mathbf{y}\in\mathcal{S}^L}$. By combining (4.12) and (4.10), $g_n(u_n, \mathbf{x}_n)$ can be simplified as follows:

$$g_n(u_n, \mathbf{x}_n) = \sum_{\mathbf{Y}} \beta_n(u_n, \mathbf{y}) f(\mathbf{x}_n|\mathbf{y}) \tag{4.13}$$

where

$$\begin{aligned}\beta_n(u_n, \mathbf{y}) &= \sum_{\mathbf{u}^n} \sum_{u_0=0}^{K-1} \pi_{u_0} f(\mathbf{y}|H_{u_0}) P(u_0|\mathbf{u}^n, u_n) \\ &\qquad \times \int_{\mathbf{X}^n} P(\mathbf{u}^n|\mathbf{x}^n) f(\mathbf{x}^n|\mathbf{y}) d\mathbf{x}^n \\ &= \sum_{\mathbf{u}^n} \sum_{u_0=0}^{K-1} \pi_{u_0} f(\mathbf{y}|H_{u_0}) P(u_0|\mathbf{u}^n, u_n) \\ &\quad \times \prod_{h=1, h\neq n}^{N} \int_{\mathbf{X}_h} P(u_h|\mathbf{x}_h) f(\mathbf{x}_h|\mathbf{y}) d\mathbf{x}_h \end{aligned} \tag{4.14}$$

is a scalar function of u_n and $\mathbf{y}$. Thus, the optimal sensor rule γ_n is

$$\gamma_n(\mathbf{x}_n) = \arg\max_{u_n} \sum_{\mathbf{Y}} \beta_n(u_n, \mathbf{y}) f(\mathbf{x}_n|\mathbf{y}). \tag{4.15}$$

For a binary hypothesis testing problem, i.e, $K = 2$, the optimal decision rule in (4.15) can be written as follows:

$$\gamma_n(\mathbf{x}_n) = \mathbb{1}\left\{ \sum_{\mathbf{Y}} [\beta_n(1, \mathbf{y}) - \beta_n(0, \mathbf{y})] f(\mathbf{x}_n|\mathbf{y}) \right\} \tag{4.16}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function defined as follows:

$$\mathbb{1}\{x\} = \begin{cases} 1, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.17}$$

Next, the optimal fusion rule at the FC that maximizes the probability of correct classification P_c is investigated. We have

$$\begin{aligned} P_c &= \sum_{u_0=0}^{K-1} P(u_0|H_{u_0})\pi_{u_0} \\ &= \sum_{\mathbf{u}} \sum_{u_0=0}^{K-1} P(u_0|\mathbf{u})P(\mathbf{u}|H_{u_0})\pi_{u_0} \end{aligned} \tag{4.18}$$

To maximize P_c, the optimal fusion rule is to make a decision u_0 such that such $\pi_{u_0}P(\mathbf{u}|H_{u_0})$ is maximized, namely

$$\begin{aligned} \gamma_0(\mathbf{u}) &= \arg\max_{u_0} \pi_{u_0}P(\mathbf{u}|H_{u_0}) \\ &= \arg\max_{u_0} \int_{\mathbf{X}} \prod_{n=1}^{N} P(u_n|\mathbf{x}_n) f(\mathbf{r}|H_{u_0})\pi_{u_0} d\mathbf{x} \end{aligned} \tag{4.19}$$

for any local decision vector $\mathbf{u}$. Because for a given $\mathbf{u}$, $P(u_0|\mathbf{u})$ takes the value either 0 or 1, the fusion rule in (4.19) maximizes P_c given in (4.18). For a binary hypothesis testing problem, the following fusion rule can be obtained:

$$\gamma_0(\mathbf{u}) = \mathbb{1}\left\{\pi_1 f(\mathbf{u}|H_1) - \pi_0 f(\mathbf{u}|H_0)\right\} \tag{4.20}$$

Proposition 4.1. *Let $\{\gamma_1(\mathbf{x}_1), \ldots, \gamma_N(\mathbf{x}_N)\}$ and $\gamma_0(\mathbf{u})$ be a set of optimal sensor decision rules and an optimal fusion rule in a distributed classification system that maximizes the probability of correct classification P_c. Then they must satisfy the following conditions:*

1) For all local decision rules $\gamma_n, n = 1, \ldots, N$:

$$\gamma_n(\mathbf{x}_n) = \arg\max_{u_n} \sum_{\mathbf{Y}} \beta_n(u_n, \mathbf{y}|\boldsymbol{\gamma}^n, \gamma_0) f(\mathbf{x}_n|\mathbf{y}) \tag{4.21}$$

where $\beta_n(\cdot)$ is given as follows:

$$\begin{aligned} & \beta_n(u_n, \mathbf{y}|\boldsymbol{\gamma}^n, \gamma_0) \\ = & \sum_{\mathbf{u}^n} \sum_{u_0=0}^{K-1} \pi_{u_0} f(\mathbf{y}|H_{u_0}) \delta\left(\gamma_0(\mathbf{u}^n, u_n) - u_0\right) \\ & \times \prod_{h=1, h\neq n}^{N} \int_{\mathbf{X}_h} \delta\left(\gamma_h(\mathbf{x}_h) - u_h\right) f(\mathbf{x}_h|\mathbf{y}) d\mathbf{x}_h \end{aligned} \tag{4.22}$$

where the notation of $\beta(\cdot|\boldsymbol{\gamma}^n, \gamma_0)$ is to emphasize that the value of β is conditioned on the given decision rules $\boldsymbol{\gamma}^n, \gamma_0$, and $\delta(\cdot)$ is defined as follows:

$$\delta(x) = \begin{cases} 1, & x = 0 \\ 0, & \textit{otherwise.} \end{cases} \tag{4.23}$$

2) For the fusion rule:

$$\gamma_0(\mathbf{u}) = \arg\max_{u_0} \int_{\mathbf{X}} \prod_{n=1}^{N} \delta\left(\gamma_n(\mathbf{x}_n) - u_n\right) f(\mathbf{x}|H_{u_0}) \pi_{u_0} d\mathbf{x}. \tag{4.24}$$

The necessary conditions to determine the optimal rules that maximize the probability of correct classification naturally is obtained. To search for the optimal rules, we adapt the idea of the Gauss-Seidel iterative algorithm. We present a computationally efficient iterative algorithm for obtaining discrete approximations of the optimal rules in the next section.

4.4 Computational Algorithm

We propose an iterative algorithm based on Proposition 4.1, by considering the following Gauss-Seidel iterative process [105]. Let the sensor decision rules and the fusion rule at the kth stage of iteration be denoted by $(\gamma_1^{(k)}, \ldots, \gamma_N^{(k)}, \gamma_0^{(k)})$ with the initial set $(\gamma_1^{(0)}, \ldots, \gamma_N^{(0)}, \gamma_0^{(0)})$. At the $k+1$th iteration, after the decision rule of sensor $n-1$, i.e. $\gamma_{n-1}^{(k+1)}$, is updated, the decision

rule of sensor n is updated according to the following equation

$$\gamma_n^{(k+1)}(\mathbf{x}_n) = \arg\max_{u_n} \sum_{\mathbf{Y}} \beta_n(u_n, \mathbf{y}|\gamma_1^{(k+1)}, \ldots, \gamma_{n-1}^{(k+1)}, \gamma_{n+1}^{(k)}, \ldots, \gamma_N^{(k)}, \gamma_0^{(k)}) f(\mathbf{x}_n|\mathbf{y}). \tag{4.25}$$

Once every sensor's decision rule is updated, the fusion rule is obtained by

$$\gamma_0^{(k+1)}(\mathbf{u}) = \arg\max_{u_0} \int_{\mathbf{X}} \prod_{n=1}^{N} \delta\left(\gamma_n^{(k+1)}(\mathbf{x}_n) - u_n\right) f(\mathbf{x}|H_{u_0})\pi_{u_0} d\mathbf{x}. \tag{4.26}$$

This algorithm involves obtaining sensor rules $(\gamma_1^{(k)}, \ldots, \gamma_N^{(k)})$ that are continuous functions. Thus, discretizing the input space, and thus the continuous functions, is necessary for obtaining a solution in practice. For illustration purposes, we present the algorithm with each sensor making its decision based on the observation at a single time instant, i.e., $L = 1$. And for notational simplicity, x_{nl} will be written as x_n by omitting the time index l. The corresponding hidden random variable Y is a scalar variable in this case. We define these functions on equally discretized grids of $\{x_{1,1}, \ldots, x_{1,t_1}, \ldots, x_{1,T_1}\}$, $\{x_{2,1}, \ldots, x_{2,t_2}, \ldots, x_{2,T_2}\}$, $\ldots$, $\{x_{N,1}, \ldots, x_{N,t_N}, \ldots, x_{N,T_N}\}$ with Δ_n being the discretization step size of x_n. The following discretized Gauss-Seidel iterative algorithm is obtained:

Step 1: Initialize N sensor rules and the fusion rule respectively, for $n = 1, \ldots, N$.

$$\gamma_n^{(0)}(x_{n,t_n}) = i \in \mathcal{K} \quad \forall t_n = 1, \ldots, T_n \tag{4.27}$$

(4.28)

$$\gamma_0^{(0)}(\mathbf{u}) = i \in \mathcal{K} \quad \forall \mathbf{u} \in \mathcal{K}^N \tag{4.29}$$

Step 2: Iteratively, update N sensor rules and the fusion rule for better system performance. The $(k+1)$th stage of the iteration is as follows:

For $t_1 = 1, \ldots, T_1$,

$$\begin{aligned}\gamma_1^{(k+1)}(x_{1,t_1}) = & \\ & \arg\max_{u_1} \sum_{Y} \beta_1(u_1, y|\gamma_2^{(k)}, \ldots, \gamma_N^{(k)}, \gamma_0^{(k)}) \\ & \times f(x_{1,t_1}|y)\end{aligned} \tag{4.30}$$

with

$$\begin{aligned}\beta_1(u_1, y|\gamma_2^{(k)}, \ldots, \gamma_N^{(k)}, \gamma_0^{(k)}) = & \\ & \sum_{\mathbf{u}^1} \sum_{u_0=0}^{K-1} \pi_{u_0} f(y|H_{u_0}) \delta\left(\gamma_0^{(k)}(\mathbf{u}^1, u_1) - u_0\right) \\ & \times \prod_{n=2}^{N} \left[\sum_{t_n=1}^{T_n} \delta\left(\gamma_n^{(k)}(x_{n,t_n}) - u_n\right) f(x_{n,t_n}|y) \Delta x_n\right]\end{aligned} \tag{4.31}$$

For $t_2 = 1, \ldots, T_2$,

$$\begin{aligned}\gamma_2^{(k+1)}(x_{2,t_2}) = & \\ & \arg\max_{u_2} \sum_{Y} \beta_2(u_2, y|\gamma_1^{(k+1)}, \gamma_3^{(k)}, \ldots, \gamma_N^{(k)}, \gamma_0^{(k)}) \\ & \times f(x_{2,t_2}|y)\end{aligned} \tag{4.32}$$

with

$$\begin{aligned}
&\beta_2(u_2, y|\gamma_1^{(k+1)}, \gamma_3^{(k)}, \ldots, \gamma_N^{(k)}, \gamma_0^{(k)}) = \\
&\quad \sum_{\mathbf{u}^2} \sum_{u_0=0}^{K-1} \pi_{u_0} f(y|H_{u_0}) \delta\left(\gamma_0^{(k)}(\mathbf{u}^2, u_2) - u_0\right) \\
&\quad \times \left[\sum_{t_1=1}^{T_1} \delta\left(\gamma_1^{(k+1)}(x_{1,t_1}) - u_1\right) f(x_{1,t_1}|y)\Delta x_1\right] \\
&\quad \times \prod_{n=3}^{N} \left[\sum_{t_n=1}^{T_n} \delta\left(\gamma_n^{(k)}(x_{n,t_n}) - u_n\right) f(x_{n,t_n}|y)\Delta x_n\right]
\end{aligned} \tag{4.33}$$

$$\vdots$$

For $t_N = 1, \ldots, T_N$,

$$\begin{aligned}
&\gamma_N^{(k+1)}(x_{N,t_N}) = \\
&\quad \arg\max_{u_N} \sum_{Y} \beta_N(u_N, y|\gamma_1^{(k+1)}, \ldots, \gamma_{N-1}^{(k+1)}, \gamma_0^{(k)}) \\
&\quad \times f(x_{N,t_N}|y)
\end{aligned} \tag{4.34}$$

with

$$\begin{aligned}
&\beta_N(u_N, y|\gamma_1^{(k+1)}, \ldots, \gamma_{N-1}^{(k+1)}, \gamma_0^{(k)}) = \\
&\quad \sum_{\mathbf{u}^N} \sum_{u_0=0}^{K-1} \pi_{u_0} f(y|H_{u_0}) \delta\left(\gamma_0^{(k)}(\mathbf{u}^N, u_N) - u_0\right) \\
&\quad \times \prod_{n=1}^{N-1} \left[\sum_{t_n=1}^{T_n} \delta\left(\gamma_n^{(k+1)}(x_{n,t_n}) - u_n\right) f(x_{n,t_n}|y)\Delta x_n\right]
\end{aligned} \tag{4.35}$$

For $\mathbf{u} \in \mathcal{K}^N$,

$$\begin{aligned}\gamma_0^{(k+1)}(\mathbf{u}) = \\ \arg\max_{u_0} \sum_{t_1}^{T_1} \cdots \sum_{t_N}^{T_N} \prod_{n=1}^{N} \delta\left(\gamma_n^{(k+1)}(x_{n,t_n}) - u_n\right) \\ \times f(x_{1,t_1}, \ldots, x_{N,t_N}|H_{u_0})\pi_{u_0}\end{aligned} \tag{4.36}$$

To guarantee the convergence of the algorithm, whenever the maximization step renders more than one values, we choose the one that matches with the previous iteration, otherwise we choose randomly.

Step 3: The termination criterion of the iteration is as follows

$$\gamma_1^{(k+1)}(\mathbf{x}_{n,t_n}) = \gamma_1^{(k)}(\mathbf{x}_{n,t_n}), \forall n = 1, \ldots, N \tag{4.37}$$

$\forall t_1, \ldots, t_n$. And for all $\forall \mathbf{u} \in \{0, \ldots, S-1\}^N$

$$\gamma_0^{(k+1)}(\mathbf{u}) = \gamma_0^{(k)}(\mathbf{u}) \tag{4.38}$$

After these decision rules $\gamma_0, \gamma_1, \ldots, \gamma_N$ are obtained, they are implemented at corresponding local sensors and FC, according to which local decisions and final decision are made.

We denote the discretized version of P_c after the k-th iteration as

$$\begin{aligned}P_c^D(\gamma_1^{(k)}, \ldots, \gamma_N^{(k)}, \gamma_0^{(k)}) = \\ \sum_{t_1=1}^{T_1} \cdots \sum_{t_N=1}^{T_N} \sum_{\mathbf{u}} \sum_{u_0=1}^{K-1} \pi_{u_0} \delta\left(\gamma_0^{(k)} - u_0\right) \\ \times \prod_{l=1}^{N} \delta\left(\gamma_n^{(k)}(\mathbf{x}_{n,t_n}) - u_n\right) f(\mathbf{x}_{1,t_1}, \ldots, \mathbf{x}_{N,t_N}|H_{u_0}) \Delta\mathbf{x}_1 \ldots \Delta\mathbf{x}_N\end{aligned} \tag{4.39}$$

Theorem 4.1. *For any positive discretization step sizes $\Delta\mathbf{x}_1, \ldots, \Delta\mathbf{x}_N$, and any initialization*

$(\gamma_1^{(0)}, \ldots, \gamma_N^{(0)}, \gamma_0^{(0)})$ *in (4.28), (4.29),* $P_c^D(\gamma_1^{(k)}, \ldots, \gamma_N^{(k)})$ *in (4.39) must converge to a local optimal value in finite number of steps.*

Proof. We denote the discretized version of P_c at the $k+1$-th iteration right after sensor n updates its rule as

$$P_c^D(\gamma_1^{(k+1)}, \ldots, \gamma_n^{(k+1)}, \gamma_{n+1}^{(k)}, \ldots, \gamma_N^{(k)}, \gamma_0^{(k)})$$

And we denote the discretized version of (4.13) as

$$\begin{aligned} g_n^D(u_n, \mathbf{x}_{n,t_n}) = & \\ & \sum_{\mathbf{Y}} \beta_n(u_n, \mathbf{y} | \gamma_1^{(k+1)}, \ldots, \gamma_{n-1}^{(k+1)}, \gamma_{n+1}^{(k)}, \ldots, \gamma_N^{(k)}, \gamma_0^{(k)}) \\ & \times f(\mathbf{x}_{n,t_n} | \mathbf{y}) \end{aligned} \tag{4.40}$$

with $\beta_n(\cdot)$ given in (4.31, 4.33, ..., 4.35). We first need to prove that the discretized probability of correct classification is nondecreasing with each single update of sensor rule or fusion rule. In other words:

1. For sensor rules: $P_c^D(\gamma_1^{(k+1)}, \ldots, \gamma_n^{(k+1)}, \gamma_{n+1}^{(k)}, \ldots, \gamma_N^{(k)}, \gamma_0^{(k)})$ is nondecreasing as n increase

2. For fusion rule:

$$\begin{aligned} & P_c^D(\gamma_1^{(k+1)}, \ldots, \gamma_N^{(k+1)}, \gamma_0^{(k)}) \\ & \quad \leq P_c^D(\gamma_1^{(k+1)}, \ldots, \gamma_N^{(k+1)}, \gamma_0^{(k+1)}) \end{aligned} \tag{4.41}$$

We first prove the nondecreasing property of P_c^D with respect to n.

$$
\begin{aligned}
&P_c^D(\gamma_1^{(k+1)},\ldots,\gamma_n^{(k+1)},\gamma_{n+1}^{(k)},\ldots,\gamma_N^{(k)}) \\
&\quad=\sum_{t_n=1}^{T_n}\sum_{u_n}\delta\left(\gamma_n^{(k+1)}(\mathbf{x}_{n,t_n})-u_n\right)\sum_{\mathbf{Y}}f(\mathbf{x}_{n,t_n}|\mathbf{y}) \\
&\quad\times\beta_n(u_n,\mathbf{y}|\gamma_1^{(k+1)},\ldots,\gamma_{n-1}^{(k+1)},\gamma_{n+1}^{(k)},\ldots,\gamma_N^{(k)},\gamma_0^{(k)}) \\
&\quad=\sum_{t_n=1}^{T_n}\sum_{u_n}\delta\left(\gamma_n^{(k+1)}(\mathbf{x}_{n,t_n})-u_n\right)g_n^D(u_n,\mathbf{x}_{n,t_n}) \\
&\quad=\sum_{t_n=1}^{T_n}\sum_{u_n}\delta\left(\gamma_n^{(k)}(\mathbf{x}_{n,t_n})-u_n\right)g_n^D(u_n,\mathbf{x}_{n,t_n}) \\
&\quad+\sum_{t_n=1}^{T_n}\sum_{u_n}\left[\delta\left(\gamma_n^{(k+1)}(\mathbf{x}_{n,t_n})-u_n\right)-\delta\left(\gamma_n^{(k)}(\mathbf{x}_{n,t_n})-u_n\right)\right] \\
&\quad\times g_n^D(u_n,\mathbf{x}_{n,t_n}) \\
&\quad=P_c^D(\gamma_1^{(k+1)},\ldots,\gamma_{n-1}^{(k+1)},\gamma_n^{(k)},\ldots,\gamma_N^{(k)}) \\
&\quad+\sum_{t_n=1}^{T_n}[\sum_{u_n}\delta\left(\gamma_n^{(k+1)}(\mathbf{x}_{n,t_n})-u_n\right)g_n^D(u_n,\mathbf{x}_{n,t_n}) \\
&\quad-\sum_{u_n}\delta\left(\gamma_n^{(k)}(\mathbf{x}_{n,t_n})-u_n\right)g_n^D(u_n,\mathbf{x}_{n,t_n})] \\
&\quad\geq P_c^D(\gamma_1^{(k+1)},\ldots,\gamma_{n-1}^{(k+1)},\gamma_n^{(k)},\ldots,\gamma_N^{(k)})
\end{aligned}
\tag{4.42}
$$

The last inequality is because according to the iterative steps in the algorithm for given $\mathbf{x}_{n,t_n}$, $\gamma_n^{(k+1)}(\mathbf{x}_{n,t_n})$ is determined by maximizing $g_n^D(u_n,\mathbf{x}_{n,t_n})$. Next, we need to prove that $P_c^D(\gamma_1^{(k+1)},\ldots,\gamma_N^{(k+1)},\gamma_0^{(k)})\leq P_c^D(\gamma_1^{(k+1)},\ldots,\gamma_N^{(k+1)},\gamma_0^{(k+1)})$. We denote the discretized version of $P(\mathbf{u}|H_i)$ at $k+1$-th iteration as:

$$
\begin{aligned}
&P(\mathbf{u}|H_i)= \\
&\quad\sum_{t_1=1}^{T_1}\cdots\sum_{t_N=1}^{T_N}\sum_{\mathbf{Y}}f(\mathbf{y}|H_i)\prod_{n=1}^{N}\delta\left(\gamma_n^{(k+1)}(\mathbf{x}_{n,t_n})-u_n\right)f(\mathbf{x}_{n,t_n}|\mathbf{y})
\end{aligned}
\tag{4.43}
$$

$$
\begin{aligned}
&P_c^D(\gamma_1^{(k+1)},\ldots,\gamma_N^{(k+1)},\gamma_0^{(k+1)}) \\
&\quad=\sum_{\mathbf{u}}\sum_{u_0=0}^{K-1}\delta\left(\gamma_0^{(k+1)}-u_0\right)\pi_{u_0}P(\mathbf{u}|H_{u_0}) \\
&\quad=\sum_{\mathbf{u}}\sum_{u_0=0}^{K-1}\delta\left(\gamma_0^{(k)}-u_0\right)\pi_{u_0}P(\mathbf{u}|H_{u_0}) \\
&\quad+\sum_{\mathbf{u}}\sum_{u_0=0}^{K-1}\left[\delta\left(\gamma_0^{(k+1)}-u_0\right)-\delta\left(\gamma_0^{(k)}-u_0\right)\right]\pi_{u_0}P(\mathbf{u}|H_{u_0}) \\
&\quad=P_c^D(\gamma_1^{(k+1)},\ldots,\gamma_N^{(k+1)},\gamma_0^{(k)}) \\
&\quad+\sum_{\mathbf{u}}[\sum_{u_0=0}^{K-1}\delta\left(\gamma_0^{(k+1)}-u_0\right)\pi_{u_0}P(\mathbf{u}|H_{u_0}) \\
&\quad-\sum_{u_0=0}^{K-1}\delta\left(\gamma_0^{(k)}-u_0\right)\pi_{u_0}P(\mathbf{u}|H_{u_0})] \\
&\quad\geq P_c^D(\gamma_1^{(k+1)},\ldots,\gamma_N^{(k+1)},\gamma_0^{(k)})
\end{aligned}
\tag{4.44}
$$

The last inequality is because at the $k+1$-th iteration, for given $\mathbf{u}$, $\gamma_0^{(k+1)}(\mathbf{u})$ is chosen to maximize $\pi_{u_0}P(\mathbf{u}|H_{u_0})$. Thus, it is proved that with each iteration, the discretized probability of correct classification is nondecreasing. Since $P_c^D(\gamma_1^{(k)},\ldots,\gamma_N^{(k)},\gamma_0^{(k)})\leq 1$ can only take finite number of values, it can be concluded that P_c^D converges to a stationary point in finite number of steps. If the limit of the algorithm is not a local/global maximum, $P_c^D(\gamma_1^{(k)},\ldots,\gamma_N^{(k)},\gamma_0^{(k)})$ could be larger than its value at the stationary point. Thus, $P_c^D(\gamma_1^{(k)},\ldots,\gamma_N^{(k)},\gamma_0^{(k)})$ converges to a local optimal solution in finite number of steps. □

Theorem 4.2. *For any positive discretization step sizes $\Delta\mathbf{x}_1,\ldots,\Delta\mathbf{x}_N$, and any initialization $(\gamma_1^{(0)},\ldots,\gamma_N^{(0)},\gamma_0^{(0)})$ in (4.28), (4.29), the algorithm terminates within a finite number of iterations.*

Proof. It has been proved that within a finite number of iterations, $P_c^D(\gamma_1^{(k)},\ldots,\gamma_N^{(k)},\gamma_0^{(k)})$ converges to a stationary point. Beyond the convergence of P_c^D, we need to prove that the algorithm cannot infinitely oscillate, namely

1. For all $n=1,\ldots,N$, $\gamma_n^{(k+1)}(\mathbf{x}_{n,t_n})=\gamma_n^{(k)}(\mathbf{x}_{n,t_n}),\forall t_n=1\ldots,T_n$ is satisfied with finite

number of iterations;

2. And for all $\forall \mathbf{u} \in \{0, \ldots, K-1\}^N$, $\gamma_0^{(k+1)}(\mathbf{u}) = \gamma_0^{(k)}(\mathbf{u})$ is satisfied within a finite number of iterations.

By Theorem 4.1, P_c^D must attain a stationary point within a finite number of iterations, i.e.,

$$P_c^D(\gamma_1^{(k+1)}, \ldots, \gamma_n^{(k+1)}, \gamma_{n+1}^{(k)}, \ldots, \gamma_N^{(k)}) = P_c^D(\gamma_1^{(k+1)}, \ldots, \gamma_n^{(k)}, \gamma_{n+1}^{(k)}, \ldots, \gamma_N^{(k)}) \tag{4.45}$$

By (4.42), we have $\forall t_n = 1, \ldots, T_n$

$$\sum_{u_n} \delta\left(\gamma_n^{(k+1)}(\mathbf{x}_{n,t_n}) - u_n\right) g_n^D(u_n, \mathbf{x}_{n,t_n}) - \sum_{u_n} \delta\left(\gamma_n^{(k)}(\mathbf{x}_{n,t_n}) - u_n\right) g_n^D(u_n, \mathbf{x}_{n,t_n}) = 0 \tag{4.46}$$

which implies either,

$$\gamma_n^{(k+1)}(\mathbf{x}_{n,t_n}) = \gamma_n^{(k)}(\mathbf{x}_{n,t_n}), \forall n = 1, \ldots, N \tag{4.47}$$

or that $g_n^D(u_n, \mathbf{x}_{n,t_n})$ is maximized at multiple values of u_n. However, if the decision of the k-th step $\gamma_n^{(k)}(\mathbf{x}_{n,t_n})$ maximizes $g_n^D(u_n, \mathbf{x}_{n,t_n})$ along with other values of u_n, then the decision at the $k+1$-th step $\gamma_n^{(k+1)}(\mathbf{x}_{n,t_n})$ remains consistent with the previous decision. Thus, the algorithm can not oscillate infinitely among several values. By (4.44), we have for any $\mathbf{u}$

$$\sum_{u_0=0}^{K-1} \delta\left(\gamma_0^{(k+1)} - u_0\right) \pi_{u_0} P(\mathbf{u}|H_{u_0}) - \sum_{u_0=0}^{K-1} \delta\left(\gamma_0^{(k)} - u_0\right) \pi_{u_0} P(\mathbf{u}|H_{u_0}) = 0, \tag{4.48}$$

which implies that

$$\gamma_0^{(k+1)}(\mathbf{u}) = \gamma_0^{(k)}(\mathbf{u}) \tag{4.49}$$

It follows that when P_c^D converges to a stationary point $\gamma_l^{(k+1)}(\mathbf{r}_{l,t_l})$ and $\gamma_0^{(k+1)}(\mathbf{u})$ are invariant. Thus, the algorithm terminates within a finite number of iterations. □

It has to be noted that Theorem 4.2 does not guarantee convergence of the algorithm to the global maximum for any initialization. Even when the discretization step approaches zero, the algorithm converges to a person-by-person optimal solution. The following theorem shows that under some conditions, when the discretization step approaches zero, the global maximum value that can be obtained by our algorithm achieves the maximum of P_c , which is defined as $P_{c\sup} := \sup_{\gamma_1,\ldots,\gamma_N,\gamma_0} P_c(\gamma_1,\ldots,\gamma_N,\gamma_0)$.

Theorem 4.3. *Let $\Delta\mathbf{x}_1 = \Delta\mathbf{x}_2 = \cdots = \Delta\mathbf{x}_N = \Delta$, and let $P_{c\Delta}$ be the global maximum of the discrete version of P_c. When Δ approaches zero, $P_{c\Delta}$ achieves the maximum of $P_{c\sup}$, i.e.,*

$$\lim_{\Delta\to 0} P_{c\Delta} = P_{c\sup} \tag{4.50}$$

if for a series of regions $\Omega_0,\ldots,\Omega_{K-1}$

$$\Omega_i = \{(\mathbf{x}_1,\ldots,\mathbf{x}_N) : \gamma_0(\gamma_1(\mathbf{x}_1),\ldots,\gamma_N(\mathbf{x}_N)) = i\}, i = 1,\ldots,K-1$$

defined by any set of decision rules and fusion rule $(\gamma_1,\ldots,\gamma_N,\gamma_0)$, the following inequality holds

$$\left|\sum_{i=1}^{K-1}\pi_i\left[\int_{\Omega_i} p(\mathbf{x}_1,\ldots,\mathbf{x}_N|H_i)d\mathbf{x}_1\ldots d\mathbf{x}_N - RS(\Omega_i,\Delta)\right]\right| \le a\Delta \tag{4.51}$$

where $RS(\Omega_i,\Delta)$ is a Riemann sum of the corresponding integral term and the constant a does

not depend on Ω_i *and* Δ.

Proof. According to the definition of $P_{c\sup}$, a set of decision rules and fusion rule $(\gamma_1, \ldots, \gamma_N, \gamma_0)$ exist such that for all $\epsilon > 0$

$$P_c(\gamma_1, \ldots, \gamma_N, \gamma_0) + \frac{1}{2}\epsilon \geq P_{c\sup} \tag{4.52}$$

The Riemann sum of $P_c(\gamma_1, \ldots, \gamma_N, \gamma_0)$ corresponding to this particular decision set $(\gamma_1, \ldots, \gamma_N, \gamma_0)$ is denoted as $P_c^{\Delta}(\gamma_1, \ldots, \gamma_N, \gamma_0)$. There exists a $\Delta^* > 0$ such that for any $\Delta < \Delta^*$

$$P_c^{\Delta}(\gamma_1, \ldots, \gamma_N, \gamma_0) + \frac{1}{2}\epsilon \geq P_c(\gamma_1, \ldots, \gamma_N, \gamma_0) \tag{4.53}$$

Thus,

$$P_c^{\Delta}(\gamma_1, \ldots, \gamma_N, \gamma_0) + \epsilon \geq P_{c\sup} \tag{4.54}$$

By the definition of $P_{c\Delta}$, we have

$$P_{c\Delta} + \epsilon \geq P_c^{\Delta}(\gamma_1, \ldots, \gamma_N, \gamma_0) + \epsilon \geq P_{c\sup}, \forall \Delta \leq \Delta^* \tag{4.55}$$

From the inequalities in (4.55), it can be concluded that

$$\liminf_{\Delta \to 0} P_{c\Delta} + \epsilon \geq P_{c\sup} \tag{4.56}$$

Since ϵ is arbitrary, the above inequality can be written as

$$\liminf_{\Delta \to 0} P_{c\Delta} \geq P_{c\sup} \tag{4.57}$$

Equation (4.50) can be proved if the following inequality can be proved.

$$\limsup_{\Delta \to 0} P_{c\Delta} \leq P_{c\sup} \tag{4.58}$$

To do that, we first assume the contrary of (4.58) to be true, namely

$$\limsup_{\Delta \to 0} P_{c\Delta} > P_{c\sup} \tag{4.59}$$

Then a positive $\delta > 0$ and a sequence $\{\Delta_z\}$ would exist such that as $\Delta_z \to 0$

$$P_{c\Delta_z} > P_{c\sup} + \delta \tag{4.60}$$

For every such Δ_z, a set of decision rules $(\gamma_1^{(z)}, \ldots, \gamma_N^{(z)}, \gamma_0^{(z)})$ must exist such that

$$P_{c\Delta_z} = P_c^{\Delta_z}(\gamma_1^{(z)}, \ldots, \gamma_N^{(z)}, \gamma_0^{(z)}) \tag{4.61}$$

By the condition in the theorem

$$P_c(\gamma_1^{(z)}, \ldots, \gamma_N^{(z)}, \gamma_0^{(z)}) \geq P_{c\Delta_z} - \delta \tag{4.62}$$

Combined with (4.60), we have

$$P_c(\gamma_1^{(z)}, \ldots, \gamma_N^{(z)}, \gamma_0^{(z)}) > P_{c\sup} \tag{4.63}$$

which contradicts the definition of $P_{c\sup}$. Therefore, the inequality in (4.58) should be true. So,

$$P_{c\sup} \leq \liminf_{\Delta \to 0} P_{c\Delta} \leq \limsup_{\Delta \to 0} P_{c\Delta} \leq P_{c\sup} \tag{4.64}$$

and thus (4.50) is proved. □

4.5 Example

In this section, we apply our proposed approach to a distributed AMC problem. For the candidate hypothesis being a M-PSK signal, the constellation symbol set is given as $\mathcal{S}_i = \{e^{j2\pi m_i/M_i} | m_i = 0, \ldots, M_i - 1\}$ while for the candidate hypothesis being a M-QAM signal, the constellation symbol set is $\mathcal{S}_i = \{b_{m_i} e^{j\theta_{m_i}} | m_i = 1, \ldots, M_i\}$, where b_{m_i} is the amplitude of the m_i-th symbol. Binary hypothesis testing is considered in our example, i.e., $K = 2$.

We assume that the wireless channel between the unknown transmitter and each sensor undergoes flat block fading, i.e., the channel impulse response is

$$h(t) = \alpha e^{j\theta} \delta(t)$$

where α and θ are the channel (or the signal) gain and the channel (or the signal) phase, respectively, which are assumed known in this work. We assume the observation noise w_{nl} to be independent and identically distributed (i.i.d) circularly symmetric complex Gaussian with real and imaginary parts of variance $N_0/2$, i.e., $w_{nl} \sim \mathcal{CN}(0, N_0)$. The PDF of x_{nl} conditioned on the modulation format i and the symbol s_l can be written as

$$f(x_{nl}|s_l, H_i) = \frac{1}{\pi N_0} \exp\left(-\frac{1}{N_0}|x_{nl} - \alpha e^{j\theta} s_l|^2\right) \tag{4.65}$$

To evaluate the performance of the decision rules obtained from the iterative algorithm for distributed AMC, denoted by IADA, we consider several binary modulation classification problems in this section. We first consider a WSN consisting of two receiving sensors ($N = 2$), for each of which, the length of the local decision window is $L = 1$.

In the initialization step, local sensor decisions $\gamma_n(x_{n,t_n})$ are chosen randomly and the

fusion rule is the majority voting rule. Through multiple experiments, we are able to observe that different initializations of the local rules eventually lead to the same set of decision rules, while different initializations of the fusion rule result in different outcomes. Thus, multiple initializations of the fusion rule are needed for comparison purposes and the majority voting rule leads to the best performance in our case. Then, sensor decision rules and the fusion rule are updated iteratively until the termination conditions in (4.37), (4.38) are satisfied. After the rules are obtained, $100,000$ Monte Carlo trials are conducted to test the performance of these rules, in terms of the probability of correct classification. Following the above procedure, the performance of the algorithm in Section 4.4 is obtained for different signal to noise ratio (SNR) values.

In the first simulation experiment, the two candidate hypotheses are both PSK and equal priors are assumed, i.e., $\pi_0 = \pi_1 = 0.5$. Many sets of simulation have been conducted for different combinations of candidate hypotheses. Due to the similarity of the results, the performance of IADA is given for only two sets of simulations (BPSK vs. QPSK and QPSK vs. 8PSK) in Figure 4.2, where it is compared with the likelihood ratio based method (LRBM) derived under the independence assumption [124]. Our proposed algorithm achieves a much better performance in distinguishing two different PSKs for the SNR values considered.

In the second simulation experiment, we test a PSK signal against a QAM signal, and the priors are set to $\pi_0 = 0.3, \pi_1 = 0.7$. The performance when testing 16QAM against BPSK and testing 16QAM against QPSK is shown in Figure 4.3, demonstrating the superiority of the decision rules obtained by our algorithm compared to the independence-assumption based method.

The previous simulations about binary modulation classification were conducted for a two-sensor network. We further test our proposed sensor rules and fusion rule in networks with multiple sensors ($N \geq 2$). As is shown in Figure 4.5, the classification performance improves with the increase in the size of sensor network.

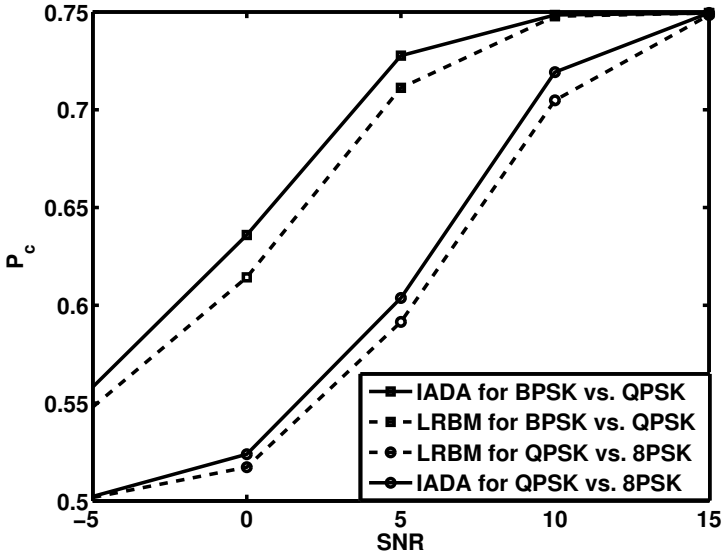

Fig. 4.2: P_c vs. SNR for testing between two PSK modulation schemes

We also conduct an experiment for a multiple hypotheses scenario. We assume that there are three candidate hypotheses, each can be a M-PSK or a M-QAM. The simulation results for classifying three hypotheses are shown in Figure 4.5. It can be seen that the decision rules obtained using our iterative algorithm perform better.

4.6 Summary

In this work, we studied the problem of distributed classification of random signals in sensor networks where sensor observations are statistically dependent due to the randomness of signals. With the introduction of a “hidden" random variable, we successfully derived the necessary conditions for the optimal sensor compression rules and fusion rule under conditionally dependent observations. An iterative algorithm, which is easy to implement, was proposed to search for the sensor rules and the fusion rule according to the derived necessary conditions. We proved that with any initialization, the proposed algorithm converges to a local optimal point within a finite number of iterations. It was also shown that when the discretization step

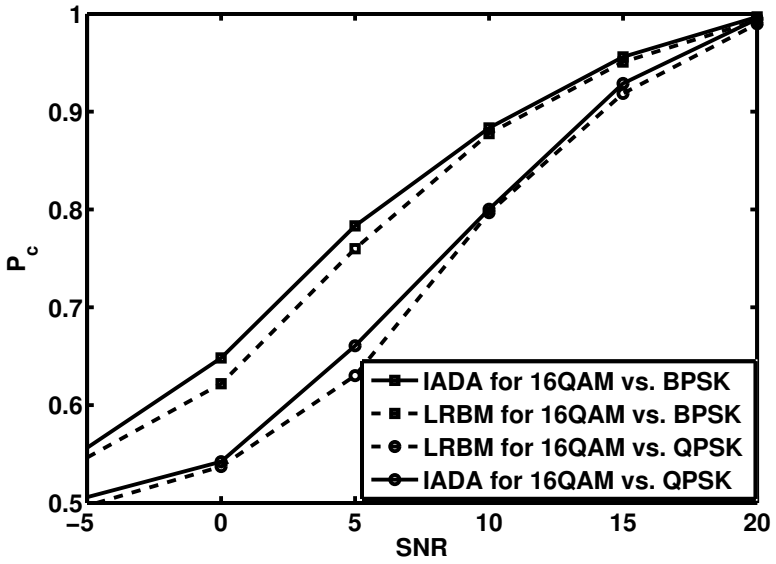

Fig. 4.3: P_c vs. SNR for testing a PSK against a QAM

size of the algorithm approaches zero, the global maximum value that can be obtained by our algorithm achieves the maximum value of P_c. As an example, we considered modulation classification problems where each modulation scheme is a candidate hypothesis. The simulation results showed that the decision rules obtained through our proposed algorithm exhibits better performance than the existing approach where independence is assumed.

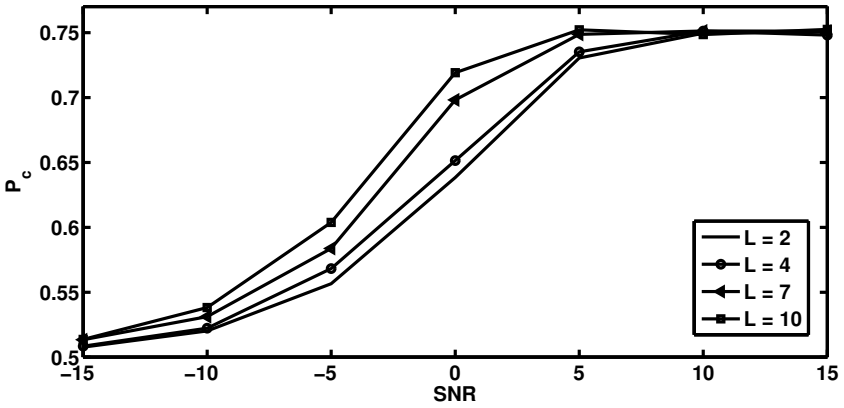

Fig. 4.4: P_c vs. SNR for testing BPSK against QPSK in sensor networks with different sizes

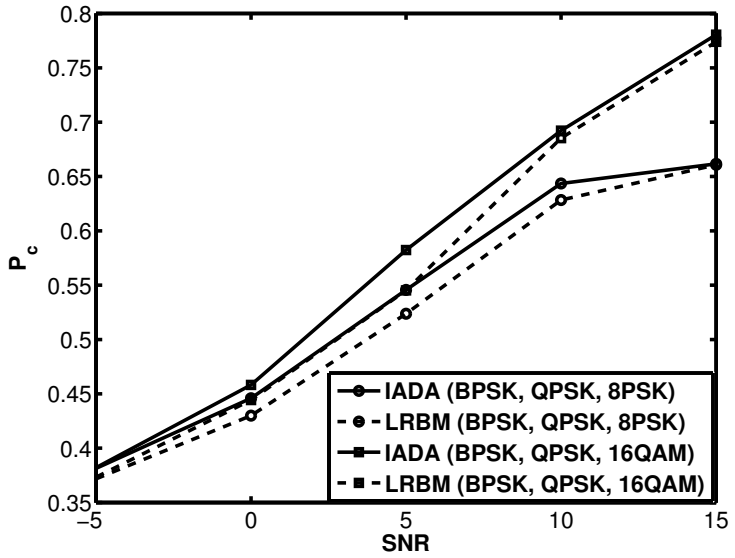

Fig. 4.5: P_c vs. SNR for 3 hypotheses testing

CHAPTER 5
FUSION OF CENSORED DEPENDENT DATA FOR DETECTION

5.1 Motivation

Advances in computational capabilities of the constituent sensor nodes inspired a surge of interest in distributed detection, in which the sensors send their locally processed data instead of raw observations to the FC, and the FC makes the final decision according to a certain fusion rule [112]. A new transmission-efficient distributed detection framework is considered in [6, 7, 55, 83], based on a send/no-send idea. The sensors “censor" their observations according to a certain mechanism to satisfy the communication rate constraints. In this process, sensors send their observations to the FC only if they are deemed “informative". Thus, only a subset of observations are received at the FC for decision making. It has been proved that with conditionally independent sensor data, transmission occurs if and only if the local likelihood ratio falls outside of a single “no-send" interval, under both Neyman-Pearson and Bayesian frameworks [7, 83].

Detection problems with censoring sensors have been investigated from various perspec-

tives, e.g., utilizing sequential detection [4], assuming fading channels [56] or under eavesdropper attacks [68]. In [106], the authors investigated the optimal censoring and transmission strategies by considering an asymptotic criterion involving error exponents in a network where sensors have access to some side information. The idea of censoring has also been applied for data reduction for estimation purposes in [74, 127, 130].

Prior research on censoring for distributed detection and estimation has been carried out under the assumption of conditionally independent observations. However, dependence often occurs in practice as the sensors observing the same phenomenon are likely to have spatially dependent observations. The design of the censoring scheme at local sensors and the fusion rule at FC becomes highly complex as a result of dependence among observations. The effect of dependence on the performance of distributed detection has been investigated recently in [2,23,50]. The authors in [59] and [120] considered physics-based models of spatial correlation and protocol-based communications and constraints, with alternative forms of censoring. The detection problem with censoring sensors considered in [3], assumes that spatial dependence among observations is known to the FC. In many practical situations, such information is not available, due to either the intrinsic non-stationarity of the signal or heterogeneity of the sensing modalities. In this chapter, we consider the fusion of censored data for distributed detection in a heterogeneous sensor network under unknown inter-sensor dependence.

5.2 Problem Formulation

We consider the detection problem in a sensor network where the two hypotheses are denoted by H_0 (null) and H_1 (target). A total of N sensors are deployed to observe the phenomenon of interest and the Fusion Center (FC) also takes its own observations. We use the random variables X_n and X_0 to respectively denote the observation of sensor n and the observation of the FC at each time instant. It is assumed that the real-valued observations of each sensor and the FC are independent and identically distributed (i.i.d) over time with known probability

density function (PDF) $f_n(x_n|H_0)$ and $f_n(x_n|H_1)$, respectively under hypotheses H_0 and H_1, i.e., for each sensor n during any time interval $1 \leq l \leq L$

$$f(x_{n1}, \ldots, x_{nL}|H_i) = \prod_{l=1}^{L} f_n(x_{nl}|H_i) \tag{5.1}$$

where x_{nl} denotes the observation at time instant l and L is the length of the decision window. However, spatial dependence exists among sensors' and FC's observations, i.e., for each time instant l,

$$f(x_{0l}, \ldots, x_{Nl}|H_i) \neq \prod_{n=0}^{N} f_n(x_{nl}|H_i) \tag{5.2}$$

and it is not specified.

In a censoring sensor network, each sensor node decides to transmit or not based on a function of its own observation $h_n(x_{nl})$, such as the likelihood ratio function. When $h_n(x_{nl})$ falls in the sending region $R_n^{'}$ the message u_{nl} is transmitted, otherwise nothing is sent. Thus, the censoring operation takes the following form:

$$\begin{cases} h_n(x_{nl}) \in R_n^{'}, & u_{nl} = {\gamma_n}^{'}(x_{nl}) \text{ is sent} \\ h_n(x_{nl}) \in \overline{R_n^{'}}, & \text{nothing is sent} \end{cases} \tag{5.3}$$

where the complement set $\overline{R_n^{'}}$ is the censoring/no-send region and ${\gamma_n}^{'}(\cdot)$ denotes the mapping from observation to message.

In many detection problems, "null" (H_0), namely the "normal condition", occurs much more frequently than "target" (H_1), and, therefore, the probability of censoring is considered only under H_0 in this chapter. In the NP framework, the censoring region $\overline{R_n^{'}}$ of each sensor satisfies an individual censoring rate constraint as follows

$$P(h_n(X_n) \in \overline{R_n^{'}}|H_0) = \beta_n \tag{5.4}$$

where $0 < \beta_n < 1$ and

$$P(h_n(X_n) \in \overline{R_n'}|H_0) = \int\limits_{h_n(x_n)\in\overline{R_n'}} f_n(x_n|H_0)dx_n \tag{5.5}$$

is the probability that sensor n censors its observations under H_0. Censoring results in the transmission of the most "informative" observations under the censoring rate constraint so as to attain the best possible detection performance.

It has been proved in [83] that given conditionally independent observations, the optimal $h_n(\cdot)$ and $\gamma_n^{'}(\cdot)$ in the censoring scheme (5.3) are both likelihood ratio functions, i.e.,

$$l_n(x_{nl}) = \frac{f_n(x_{nl}|H_1)}{f_n(x_{nl}|H_0)} \tag{5.6}$$

and $\overline{R_n'}$ is a single interval. That is only *extremal* or very "informative" likelihood ratios are transmitted. For the case of dependent observations, we assume that the censoring scheme in (5.3) is applied with $h_n(\cdot)$ being the likelihood ratio function and $\overline{R_n'}$ being a single-interval[1]. The censoring scheme can be rewritten as

$$\begin{cases} x_{nl} \in R_n, & u_{nl} = \gamma_n(x_{nl}) \text{ is sent} \\ x_{nl} \in \overline{R_n}, & \text{nothing is sent} \end{cases} \tag{5.7}$$

where

$$R_n = h_n^{-1}(R_n^{'}) \text{ and } \overline{R_n} = h_n^{-1}(\overline{R_n'}) \tag{5.8}$$

If the ratio of the two PDFs is a non-decreasing function in the argument x_{nl}, we say that they exhibit the Monotone Likelihood Ratio (MLR) property in x_{nl}. For the distributions $f_n(\cdot|H_1)$

[1]Finding the optimal censoring region in the case of dependent sensor observations is quite difficult, even with the arguably simplest case of multivariate Gaussian observations [123]. In this thesis, we continue to assume that $\overline{R_n'}$ is a single interval, even through it may not be optimal.

and $f_n(\cdot|H_0)$ satisfying the MLR proerty, i.e., $h_n(\cdot)$ is non-decreasing, it can be proved that $\overline{R_n}$ preserves the single interval nature of the censoring region according to its definition in (5.8). In this chapter, we assume that the observation at each sensor node satisfies the MLR property [2]. Thus, we have $\overline{R}_n := [t_{n1}, t_{n2}]$, where t_{n1} and t_{n2} are respectively the lower and upper limits of the no-send interval.

We further define two sets: $\mathbb{C}_l := \{n : x_{nl} \in \overline{R}_n\}$ and $\mathbb{S}_l := \{n : x_{nl} \in R_n\}$ to respectively represent the set of sensors whose observations are censored and the set of sensors whose observations are transmitted at time instant l. We use $\mathbf{u}_{\mathbb{S}_l} = \{u_{nl} : n \in \mathbb{S}_l\}$ to denote the set of messages that are transmitted from the sensors to the FC at time instant l.

Assuming ideal sensor-to-FC channels, the FC makes the decision about the true state of nature by combining the messages $\mathbf{u}_{\mathbb{S}} = \{\mathbf{u}_{\mathbb{S}_1}, \ldots, \mathbf{u}_{\mathbb{S}_L}\}$ from the sensors with its own observations $\mathbf{x}_0 = \{x_{01}, \ldots, x_{0L}\}$. We focus on designing the fusion rule $\gamma_0(\mathbf{u}_{\mathbb{S}}, \mathbf{x}_0)$ in the NP framework, assuming that each sensor's censoring scheme is known to the FC. The optimal fusion rule in the NP sense maximizes the probability of detection P_D subject to the constraint that P_F is no greater than α, i.e.,

$$\begin{aligned} &\max_{\gamma_0} P_D(\gamma_0), \\ &\text{subject to} \quad P_F(\gamma_0) \leq \alpha \end{aligned} \tag{5.9}$$

where $P_D = P(\gamma_0(\mathbf{u}_{\mathbb{S}}, \mathbf{x}_0) = 1|H_1)$ and $P_F = P(\gamma_0(\mathbf{u}_{\mathbb{S}}, \mathbf{x}_0) = 1|H_0)$.

The design of the fusion rule is considered under two transmission scenarios in which, depending on the mapping from uncensored observations to messages, either continuous or discrete data from sensors is transmitted.

Scenario A-C: Analog censored data is transmitted. Uncensored raw observations are directly transmitted to the FC, i.e., $\gamma_n(x_{nl}) = x_{nl}$ for $n \in \mathbb{S}_l$. In this case, $\mathbf{x}_{\mathbb{S}_l} = \{x_{nl} : n \in \mathbb{S}_l\}$ is received at time instant l, and the fusion of the analog censored messages $\mathbf{x}_{\mathbb{S}} =$

[2] Many families of distributions satisfy the MLR property, such as the one-parameter exponential family.

$\{\mathbf{x}_{\mathbb{S}_1}, \ldots, \mathbf{x}_{\mathbb{S}_L}\}$ along with $\mathbf{x}_0$ is considered.

Scenario Q-C: Quantized-censored data is transmitted. Uncensored observations are first quantized by a multilevel finite-range [3] uniform quantizer and then transmitted. Data that fall in the two send-zones $(-\infty, t_{n1})$ and $(t_{n2}, +\infty)$ are respectively quantized by a uniform quantizer with step size q_n that is determined by the number of bits that can be transmitted over the channel. We consider finite-range quantization with negligible saturation error. Any input signal occurring within a given quantization partition is reported at the quantizer output as being at the center of that partition (i.e., the input is rounded-off to the center of the partition). The explicit quantizer output $u_{nl} = \gamma_n(x_{nl})$ is given by:

$$\gamma_n(x_{nl}) = \begin{cases} t_{n1} - L_n q_n + q_n/2, & x_{nl} \in (-\infty, t_{n1} - (L_n - 1)q_n) \\ t_{n1} + q_n \lfloor \frac{x_{nl} - t_{n1}}{q_n} \rfloor + q_n/2, & x_{nl} \in [t_{n1} - (L_n - 1)q_n, t_{n1}) \\ t_{n2} + q_n \lfloor \frac{x_{nl} - t_{n2}}{q_n} \rfloor + q_n/2, & x_{nl} \in (t_{n2}, t_{n2} + (U_n - 1)q_n] \\ t_{n2} + U_n q_n - q_n/2, & x_{nl} \in (t_{n2} + (U_n - 1)q_n, +\infty) \end{cases} \tag{5.10}$$

where L_n and U_n are respectively the number of quantization levels of the two send-zones and $\lfloor x \rfloor$ represents the largest integer that is no greater than x. Each quantization partition can be represented by an integer $i_n \in \{-L_n, \ldots, -1, 0, 1, \ldots, U_n - 1\}$, and corresponds to the value of $k_n(i_n)$ at the quantizer output

$$k_n(i_n) = \begin{cases} t_{n1} + i_n q_n + q_n/2, & i_n < 0 \\ t_{n2} + i_n q_n + q_n/2, & i_n \geq 0 \end{cases} \tag{5.11}$$

Thus, the i_n-th quantization partition is the set of inputs corresponding to the output value $k_n(i_n)$, i.e., $Q_{i_n} := \{x_{nl} : \gamma_n(x_{nl}) = k_n(i_n)\}$. In other words, the reception of $u_{nl} = k_n(i_n)$

[3] In a finite-range quantizer, the input signals that exceed the dynamic range of the quantizer take on the value of the saturation level. The quantizers in this chapter all refer to finite range quantizers, for simplicity of presentation we refer to them only as quantizers.

indicates that the raw observation x_{nl} is in partition Q_{i_n}.

Fusion rules under these two transmission scenarios will be derived. The design of the fusion rule needs to take into consideration not only the unknown inter-sensor dependence, but also the mechanism of missing data which is the known censoring scheme to achieve better detection performance.

5.3 Copula-based Fusion

In this section, we develop the fusion rules based on copula theory for analog censored observations and quantized-censored observations respectively.

5.3.1 Fusion of Analog Censored Data

Under **Scenario A-C**, the joint PDF of received messages $\mathbf{x}_{\mathbb{S}_l}$ and FC's observation x_{0l} under hypothesis $H_i, (i = 0, 1)$ is given as

$$f(\mathbf{x}_{\mathbb{S}_l}, x_{0l}|H_i) = \int\limits_{\prod\limits_{n\in\mathbb{C}_l} \overline{R}_n} f_{\mathbf{X}}(\mathbf{x}_l, x_{0l}|H_i) d\,\mathbf{x}_{\mathbb{C}_l} \tag{5.12}$$

where $\mathbf{x}_l = \{x_{1l}, \ldots, x_{Nl}\}$, $\mathbf{x}_{\mathbb{C}_l} = \{x_{nl} : n \in \mathbb{C}_l\}$ and $f_{\mathbf{X}}(\cdot|H_i)$ denotes the joint density function of all observations $\mathbf{X} := [X_1, \ldots, X_N, X_0]$ from the sensors and the FC under H_i. We use $\prod_{n\in\mathbb{C}_l} \overline{R}_n$ to represent the multifold integration regions $\overline{R}_{\mathbb{C}_l\{1\}} \times \cdots \times \overline{R}_{\mathbb{C}_l\{|\mathbb{C}_l|\}}$ where $\mathbb{C}_l\{j\}$ is the j-th element of $\mathbb{C}_l$ and $|\cdot|$ denotes the cardinality of the set. The expression on the right hand side (RHS) of Equation (5.12) is the joint PDF of all the observations integrated over the no-send regions of all censoring sensors, which yield the joint PDF of $\{\mathbf{x}_{\mathbb{S}_l}, x_{0l}\}$ under H_i. The dimension of the integration is $|\mathbb{C}_l|$.

The unknown joint distribution $f_{\mathbf{X}}(\mathbf{x}_l, x_{0l}|H_i)$ can be approximated using a copula density function. According to the copula-based formulation of a joint distribution in (1.2), the density

function in (5.12) can be approximated by

$$\begin{aligned}
&\hat{f}(\mathbf{x}_{\mathbb{S}_l}, x_{0l}|c_i, \boldsymbol{\phi}_i, H_i) \\
&\quad = \int\limits_{\prod\limits_{n\in\mathbb{C}_l} \overline{R}_n} \hat{f}_{\mathbf{X}}(\mathbf{x}_l, x_{0l}|H_i) d\,\mathbf{x}_{\mathbb{C}_l} \\
&\quad = \int\limits_{\prod\limits_{n\in\mathbb{C}_l} \overline{R}_n} \prod_{n=0}^{N} f_n(x_{nl}|H_i) c_i(\mathbf{x}_l, x_{0l}|\boldsymbol{\phi}_i) d\,\mathbf{x}_{\mathbb{C}_l}
\end{aligned} \tag{5.13}$$

with

$$c_i(\mathbf{x}_l, x_{0l}|\boldsymbol{\phi}_i) = c_i(F_0(x_{0l}|H_i), \ldots, F_N(x_{Nl}|H_i)|\boldsymbol{\phi}_i) \tag{5.14}$$

where c_i denotes the copula density function applied to approximate the dependence structure under hypothesis H_i and $\boldsymbol{\phi}_i$ represents the corresponding dependence parameter. We have used the notation $\hat{f}(\cdot)$ to emphasize that these are approximations. How c_i is selected from a library of copula density function $\mathcal{C}$ and how the parameter $\boldsymbol{\phi}_i$ is estimated according to the data that is available at the FC, will be discussed later in the section.

The Generalized Likelihood Ratio Test (GLRT) at the FC can be obtained based on (5.13)

$$T(\mathbf{x}_{\mathbb{S}}, \mathbf{x}_0) \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{5.15}$$

where η is the threshold that satisfies the constraint $P_F(\eta) = \alpha$ and the test statistic $T(\mathbf{x}_{\mathbb{S}}, \mathbf{x}_0)$ is given as follows

$$T(\mathbf{x}_{\mathbb{S}}, \mathbf{x}_0) = \frac{\max\limits_{c_1\in\mathcal{C}, \boldsymbol{\phi}_1} \prod\limits_{l=1}^{L} \hat{f}(\mathbf{x}_{\mathbb{S}_l}, x_{0l}|c_1, \boldsymbol{\phi}_1, H_1)}{\max\limits_{c_0\in\mathcal{C}, \boldsymbol{\phi}_0} \prod\limits_{l=1}^{L} \hat{f}(\mathbf{x}_{\mathbb{S}_l}, x_{0l}|c_0, \boldsymbol{\phi}_0, H_0)} \tag{5.16}$$

It should be noted that copula selection and parameter estimation are embedded in the GLRT. The best copula c_i^* is the one that has the highest likelihood score, i.e.,

$$c_i^* = \arg\max_{c_i \in \mathcal{C}} \prod_{l=1}^{L} \hat{f}(\mathbf{x}_{\mathbb{S}_l}, x_{0l} | c_i, \hat{\boldsymbol{\phi}}_i, H_i) \tag{5.17}$$

where, for any $c_i \in \mathcal{C}$, the corresponding parameter is estimated using maximum likelihood estimation

$$\hat{\boldsymbol{\phi}}_i = \arg\max_{\boldsymbol{\phi}_i} \prod_{l=1}^{L} \hat{f}(\mathbf{x}_{\mathbb{S}_l}, x_{0l} | c_i, \boldsymbol{\phi}_i, H_i) \tag{5.18}$$

Since the true dependence model can be very complex and may not be present in the library of candidate copulas $\mathcal{C}$, the best copula c_i^* may still be misspecified.

When independent observations are assumed across the sensors, the dependence structures under both hypotheses are described by the product copula, i.e., $c_i^* = 1, i = 0, 1$. In this case, the test statistic in (5.16) reduces to

$$T(\mathbf{x}_{\mathbb{S}}, \mathbf{x}_0) = \prod_{l=1}^{L} \left[\prod_{n \in \mathbb{C}_l} \rho_n \prod_{n \in \mathbb{S}_l \cup \{0\}} \frac{f_n(x_{nl}|H_1)}{f_n(x_{nl}|H_0)} \right] \tag{5.19}$$

where ρ_n is the likelihood ratio between the two hypotheses when no message is received from sensor n, which is given as

$$\rho_n \;\; = \;\; \frac{P(X_n \in \overline{R}_n | H_1)}{P(X_n \in \overline{R}_n | H_0)} \tag{5.20}$$

The test statistic in (5.19) is the same as the one derived under independence assumption in [7, 83].

Unlike the evaluation of the test statistic in (5.19), which involves only one-dimensional integrals, the computation of $T(\mathbf{x}_{\mathbb{S}}, \mathbf{x}_0)$ in (5.16) for dependent observations involves multiple

$|\mathbb{C}_l|$-dimensional integrations due to the existence of spatial dependence. When the probability of censoring becomes higher for each sensor or the number of sensors in the network gets larger, $|\mathbb{C}_l|$ increases, so does the computational complexity of (5.16).

5.3.2 Fusion of Quantized-Censored Data

In **Scenario Q-C**, where uncensored observations are quantized before transmission, discrete-valued messages are received at the FC. A copula-based rule for fusing these discrete-valued messages and continuous observations of the FC is developed in this subsection.

Knowing local sensors' censoring schemes, the joint likelihood that the dataset $\mathbf{u}_{\mathbb{S}_l} = \{k_n(i_n) : n \in \mathbb{S}_l\}$ is received and x_{0l} is observed at the FC under hypothesis H_i is

$$f(\{k_n(i_n) : n \in \mathbb{S}_l\}, x_{0l}|H_i) = \int\limits_{\prod\limits_{n\in\mathbb{C}_l} \overline{R}_n} \int\limits_{\prod\limits_{n\in\mathbb{S}_l} Q_{i_n}} f_{\mathbf{X}}(\mathbf{x}_l, x_{0l}|H_i) d\,\mathbf{x}_l \tag{5.21}$$

where $d\,\mathbf{x}_l = dx_{1l} \ldots dx_{Nl}$ and recall that Q_{i_n} is the quantization partition corresponds to the output value $k_n(i_n)$. Eq. (5.21) is the joint distribution of sensors' and FC's observations integrated over the no-send regions of the censoring sensors and the quantization partitions Q_{i_n} of the transmitting sensors. The unknown joint distribution $f_{\mathbf{X}}(\mathbf{x}_l, x_{0l}|H_i)$ can be approximated using a copula density function. Thus, the probability density function in (5.21) can be

approximated as follows

$$\begin{aligned}
&\hat{f}(\{k_n(i_n) : n \in \mathbb{S}_l\}, x_{0l}|H_i) \\
&\quad = \int\limits_{\prod\limits_{n\in\mathbb{C}_l}\overline{R}_n} \int\limits_{\prod\limits_{n\in\mathbb{S}_l} Q_{i_n}} \hat{f}_{\mathbf{X}}(\mathbf{x}_l, x_{0l}|H_i) d\,\mathbf{x}_l \\
&\quad = \int\limits_{\prod\limits_{n\in\mathbb{C}_l}\overline{R}_n} \int\limits_{\prod\limits_{n\in\mathbb{S}_l} Q_{i_n}} \prod_{n=0}^{N} f_n(x_{nl}|H_i) c_i(\mathbf{x}_l, x_{0l}|\boldsymbol{\phi}_i) d\,\mathbf{x}_l
\end{aligned} \tag{5.22}$$

The dependence of the LHS of (5.22) on the copula model c_i and its parameter $\boldsymbol{\phi}_i$ is not specified only for notational simplicity. Statistical theory of copulas is not applied directly to approximate the joint distribution of the discrete random variables $\{u_1, \ldots, u_N\}$ because copulas for discrete marginals are not well defined. Thus, we can only approximate the joint distribution of the continuous random vector $\mathbf{X}$, through which the approximated probability of $\{\mathbf{u}_{\mathbb{S}_l}, x_{0l}\}$ under each hypothesis can be obtained. It has to be noted that an N-dimensional integration is involved in (5.22).

Based on (5.22), the joint distribution of $\mathbf{u}_{\mathbb{S}_l}$ and x_{0l} given H_i can be written as

$$\begin{aligned}
&\hat{f}(\mathbf{u}_{\mathbb{S}_l}, x_{0l}|c_i, \boldsymbol{\phi}_i, H_i) = \\
&\prod_{\mathbb{S}\in\mathcal{N}} \prod_{\substack{i_n=-L_n \\ n\in\mathbb{S}_l}}^{U_n-1} \\
&\hat{f}(\{k_n(i_n), n \in \mathbb{S}_l\}, x_{0l}|H_i)^{\mathbb{I}_{\{\mathbb{S}_l=\mathbb{S}\}}\mathbb{I}_{\left\{\mathbf{u}_{\mathbb{S}_l}=\{k_n(i_n):n\in\mathbb{S}_l\}\right\}}}
\end{aligned} \tag{5.23}$$

where $\mathbb{I}_{\{.\}}$ denotes the indicator function. $\mathbb{S}$ represents a subset of $\{1, \ldots, N\}$ and $\mathcal{N}$ represents the set consisting of all possible $\mathbb{S}$.

The Generalized Likelihood Ratio Test (GLRT) at the FC can be written as

$$T(\mathbf{u}_{\mathbb{S}}, \mathbf{x}_0) \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{5.24}$$

where η satisfies the constraint that $P_F(\eta) = \alpha$ and the test statistic $T(\mathbf{u}_{\mathbb{S}}, \mathbf{x}_0)$ is given as follows

$$T(\mathbf{u}_{\mathbb{S}}, \mathbf{x}_0) = \frac{\max\limits_{c_1 \in \mathcal{C}, \boldsymbol{\phi}_1} \prod\limits_{l=1}^{L} \hat{f}(\mathbf{u}_{\mathbb{S}_l}, \mathbf{x}_0 \,|c_1, \boldsymbol{\phi_1}, H_1)}{\max\limits_{c_0 \in \mathcal{C}, \boldsymbol{\phi}_0} \prod\limits_{l=1}^{L} \hat{f}(\mathbf{u}_{\mathbb{S}_l}, \mathbf{x}_0 \,|c_0, \boldsymbol{\phi}_0, H_0)} \tag{5.25}$$

It is noted that evaluation of $T(\mathbf{u}_{\mathbb{S}}, \mathbf{x}_0)$ involves N-dimensional integrations, thus the computational complexity increases drastically in the number of sensors. Therefore, we propose computationally efficient approximate fusion rules for both transmission scenarios in the following sections.

5.4 Noise-aided Fusion of Analog Censored Data

An alternative fusion rule for analog censored dependent data is proposed in this section based on substituting unreceived messages with artificial noise. This approach eliminates the necessity of computing multidimensional integrals and is, thus, more computationally efficient at the expense of slight performance loss.

If the FC receives no signal from sensor n at l, then we only know that $x_{nl} \in \overline{R}_n$, since neither the true underlying hypothesis nor the priors of the hypotheses are known. Thus, the uninformative prior in (5.26) is assumed for the distribution of the missing messages.

$$f(x_{nl}) = \begin{cases} \frac{1}{t_{n2} - t_{n1}}, & x_{nl} \in [t_{n1}, t_{n2}] \\ 0, & \text{otherwise} \end{cases} \tag{5.26}$$

for all $n \in \mathbb{C}_l$. An artificial noise d_{nl} is generated according to the PDF in (5.26) to represent the unreceived message from sensor n at time instant l. Let Z_n denote the message corresponding to sensor n after the addition of noise, if there is any, whose distribution under hypothesis H_i is given as

$$\begin{aligned} & f_{Z_n}(z_n|H_i) \\ & \quad = \mathbb{I}_{z_n \in R_n} f_n(z_n|H_i) + \mathbb{I}_{z_n \in \overline{R}_n} \tfrac{P(X_n \in \overline{R}_n|H_i)}{t_{n2}-t_{n1}} \end{aligned} \tag{5.27}$$

for all $n = 1, \ldots, N$. The new set of data consists of the received messages and the generated artificial noise terms, i.e., $\mathbf{z}_l := \{\mathbf{x}_{\mathbb{S}_l}, \mathbf{d}_{\mathbb{C}_l}\}$, where $\mathbf{d}_{\mathbb{C}_l} = \{d_{nl} : n \in \mathbb{C}_l\}$.

The joint PDF of the data set $\mathbf{z}_l$ and x_{0l} can be approximated using a copula density function

$$\begin{aligned} & \hat{f}_{\mathbf{Z},X_0}(\mathbf{z}_l, x_{0l}|c_i, \boldsymbol{\phi}_i, H_i) \\ & \quad = \prod_{n=1}^{N} f_{Z_n}(z_{nl}|H_i) f_0(x_{0l}|H_i) c_i(\mathbf{z}_l, x_{0l}|\boldsymbol{\phi}_i) \\ & \quad = \prod_{n \in \mathbb{C}_l} \frac{P(X_n \in \overline{R}_n|H_i)}{t_{n2}-t_{n1}} \prod_{n \in \mathbb{S}_l} f_{Z_n}(z_{nl}|H_i) f_0(x_{0l}|H_i) \\ & \quad \times c_i(\mathbf{z}_l, x_{0l}|\boldsymbol{\phi}_i) \end{aligned} \tag{5.28}$$

where

$$\begin{aligned} & c_i(\mathbf{z}_l, x_{0l}|\boldsymbol{\phi}_i) = \\ & \quad c_i(F_{Z_1}(z_{1l}|H_i), \ldots, F_{Z_N}(z_{Nl}|H_i), F_0(x_{0l}|H_i)|\boldsymbol{\phi}_i) \end{aligned} \tag{5.29}$$

Thus, the test statistic can be written as

$$T(\mathbf{z}, \mathbf{x}_0) = \frac{\max\limits_{c_1 \in \mathcal{C}, \boldsymbol{\phi}_1} \prod\limits_{l=1}^{L} \hat{f}_{\mathbf{Z},X_0}(\mathbf{z}_l, x_{0l}|c_1, \boldsymbol{\phi}_1, H_1)}{\max\limits_{c_0 \in \mathcal{C}, \boldsymbol{\phi}_0} \prod\limits_{l=1}^{L} \hat{f}_{\mathbf{Z},X_0}(\mathbf{z}_l, x_{0l}|c_0, \boldsymbol{\phi}_0, H_0)} \tag{5.30}$$

where $\mathbf{z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_L\}$. By substituting (5.28) into the above test, we have

$$\begin{aligned} T(\mathbf{z}, \mathbf{x}_0) &= \prod_{l=1}^{L} \left[\prod_{n \in \mathbb{C}_l} \rho_n \prod_{n \in \mathbb{S}_l} \frac{f_{Z_n}(z_{nl}|H_1)}{f_{Z_n}(z_{nl}|H_0)} \frac{f_0(x_{0l}|H_1)}{f_0(x_{0l}|H_0)} \right] \\ &\quad \times \frac{\max\limits_{c_1 \in \mathcal{C}, \boldsymbol{\phi}_1} \prod\limits_{l=1}^{L} c_1(\mathbf{z}_l, x_{0l}|\boldsymbol{\phi}_1)}{\max\limits_{c_0 \in \mathcal{C}, \boldsymbol{\phi}_0} \prod\limits_{l=1}^{L} c_0(\mathbf{z}_l, x_{0l}|\boldsymbol{\phi}_0)} \end{aligned} \tag{5.31}$$

The first term of the test statistic in (5.31) which is exactly the same as the test statistic under the independence assumption in (5.19), corresponds to the differences in the marginal statistics, while the spatial dependence and interactions are included in the second term.

The test in (5.31) does not require the computation of multidimensional integrals as the one in (5.16), resulting in great computational efficiency. Since the SNR at the FC is decreased due to the addition of artificial noise, detection performance is degraded, but only by a relatively small amount as will be shown later in the simulations.

5.5 Noise-aided Fusion of Quantized-Censored Data

In this section, a computationally efficient fusion rule for quantized-censored data is proposed based on Widrow's Theorem of quantization [122]. We first briefly introduce Widrow's Theorem of quantization.

5.5.1 A Review of Widrow's Statistical Theorem of Quantization

According to Widrow [121, 122], quantization of a random variable can be interpreted as the sampling of its PDF. Also, the PDF of the quantized random variable is the convolution of the original PDF with the PDF of a uniformly distributed random variable, followed by sampling.

The PDF of the uniform quantizer output u_n can be expressed as:

$$f_{U_n}(x) = (f_{W_n}(x) * f_{X_n}(x)) \sum_{t\in\mathbb{Z}} q_n \delta(x - tq_n - \frac{q_n}{2}) \tag{5.32}$$

where $*$ represents the convolution operation and $\delta(x)$ is define as

$$\delta(x) = \begin{cases} 1, & x = 0 \\ 0, & \text{otherwise} \end{cases} \tag{5.33}$$

and $f_{W_n}(x)$ is a uniform PDF as follows

$$f_{W_n}(x) = \begin{cases} \frac{1}{q_n}, & -\frac{q_n}{2} \leq x \leq \frac{q_n}{2} \\ 0, & \text{otherwise} \end{cases} \tag{5.34}$$

Uniform quantization introduces two kinds of noise: (a) the additive noise W_n and (b) aliasing error due to sampling. However, if the Characteristic Function (CF) of the input PDF $\varphi_{X_n}(\upsilon) = \mathbb{E}[e^{j\upsilon x_n}]$ is band-limited such that $\varphi_{X_n}(\upsilon) = 0$ for $|\upsilon| > \frac{\pi}{q_n}$, then in principle the original PDF of the input can be reconstructed from the knowledge of $f_{U_n}(x)$.

Theorem 5.1 (Thm QT1, [122])**.** *If the CF of X_n is "band-limited", so that*

$$\varphi_{X_n}(\upsilon) = 0, \quad |\upsilon| > \frac{\pi}{q_n} \tag{5.35}$$

then the PDF of X_n can be derived from the PDF of U_n.

Theorem 5.2 (Thm QT2, [122])**.** *If the CF of X_n is "band-limited", so that*

$$\varphi_{X_n}(\upsilon) = 0, \quad |\upsilon| > \frac{2\pi}{q_n} - \varepsilon \tag{5.36}$$

with ε positive and arbitrarily small, then the moments of X_n can be calculated from the

moments of U_n.

Noting that the conditions of Theorem 5.1 or Theorem 5.2 are more likely to be satisfied for small quantization step sizes q_n, we consider the fusion rule that is suited for high-rate quantization.

5.5.2 Computationally Efficient Fusion of Quantized-Censored Data

As discussed previously, the high complexity in computing the copula-based GLRT stems from the need for computing multidimensional integrals. We simplify the fusion process by adding controlled noise to the multilevel decisions received at the fusion center based on Widrow's theory.

Given the quantization step size q_n which is determined by the number of bits that can be transmitted over the channel. We first propose a fusion rule that corresponds to a specific censoring region $\overline{R}_n = [0, q_n]$. Then, the fusion rule is generalized to the case of any arbitrary censoring interval.

For this specific censoring interval, receiving no signal from sensor n implies that the observation of sensor n is in the quantization partition $[0, q_n]$. We can reformulate the problem as the one in which each sensor's raw observations are quantized according to the following uniform quantizer and all local quantizer outputs $\{u_{nl} : n = 1, \ldots, N\}$ are transmitted to the FC for decision making.

$$\gamma_n(x_{nl}) = \begin{cases} -L_n q_n + q_n/2, & x_{nl} < -(L_n - 1)q_n \\ q_n \lfloor x_{nl}/q_n \rfloor + q_n/2, & -(L_n - 1)q_n \leq x_{nl} < U_n q_n \\ U_n q_n + q_n/2, & x_{nl} \geq U_n q_n \end{cases} \tag{5.37}$$

According to Widrow's Theorem (5.32), the CF of uniformly quantized data contains repeated and phase-shifted replicas due to the sampling process. We are able to keep the main

lobe and filter out the terms due to aliasing in $\varphi_{U_n}(\upsilon)$ using a low pass filter (LPF). To do that, an externally generated noise d_{nl} with PDF $f_{D_n}(\cdot)$ is added to the discrete-valued observation before fusion. Let d_n denote the generated noise and the new observation z_{nl} is

$$z_{nl} = u_{nl} + d_{nl} \tag{5.38}$$

Correspondingly, the CF of the new observation is

$$\varphi_{Z_n}(\upsilon) = \varphi_{U_n}(\upsilon) \cdot \varphi_{D_n}(\upsilon) \tag{5.39}$$

The CF of noise D_n should be band-limited to play the role of a LPF. A perfect LPF-noise would have a rectangular CF in $-\frac{\pi}{q_n} \leq \upsilon \leq \frac{\pi}{q_n}$ which does not correspond to a valid PDF. Thus, attention needs to be paid while designing D_n such that as little distortion as possible is introduced when transforming the discrete-valued U_n to a continuous variable Z_n.

Most distributions that are employed in practice, like the Gaussian, exponential or chi-squared are not perfectly band-limited. This fact does not prevent the application of the quantization theorems if the quantum step size is significantly smaller than the standard deviation. If the condition stated in Theorem 5.1 is satisfied, we have the following relationship:

$$Z_n = X_n + W_n + D_n \tag{5.40}$$

Thus, at the FC, the PDF of data z_{nl} under hypothesis H_i can be derived, which is

$$f_{Z_n}(z_{nl}|H_i) = f_n(z_{nl}|H_i) * f_{W_n}(z_{nl}) * f_{D_n}(z_{nl}) \tag{5.41}$$

In practice, the quantizer output corresponding to the censoring interval $[0, q_n]$, which is $q_n/2$, is generated at the FC to represent the missing messages before the addition of LPF-noise. So, d_{nl} is added to u_{nl} for all $n \in \mathbb{S}_l$ and $d_{nl} + q_n/2$ is generated for all $n \in \mathbb{C}_l$ to

obtain the new observation z_{nl}. The joint PDF of the data $\mathbf{z}_l = \{z_{1l}, \ldots, z_{Nl}\}$ and x_{0l}, which are continuous, can be directly approximated by copula theory as follows

$$\hat{f}_{\mathbf{Z},X_0}(\mathbf{z}_l, x_{0l}|c_i, \boldsymbol{\phi}_i, H_i) = \prod_{n=1}^{N} f_{Z_n}(z_{nl}|H_i) f_0(x_{0l}|H_i) c_i(\mathbf{z}_l, x_{0l}|\boldsymbol{\phi}_i) \tag{5.42}$$

Thus, the test statistic can be written as

$$T(\mathbf{z}, \mathbf{x}_0) = \frac{\max\limits_{c_1 \in \mathcal{C}, \boldsymbol{\phi}_1} \prod\limits_{l=1}^{L} \hat{f}_{\mathbf{Z},X_0}(\mathbf{z_l}, x_{0l}|c_1, \boldsymbol{\phi}_1, H_1)}{\max\limits_{c_0 \in \mathcal{C}, \boldsymbol{\phi}_0} \prod\limits_{l=1}^{L} \hat{f}_{\mathbf{Z},X_0}(\mathbf{z_l}, x_{0l}|c_0, \boldsymbol{\phi}_0, H_0)} \tag{5.43}$$

where $\mathbf{z} = \{\mathbf{z_1}, \ldots, \mathbf{z_L}\}$. The proposed test in (5.43) is a function of continuous variables only and involve the computation of one-dimensional integrals. Compared with the test statistic in (5.23) which requires the computation of N-dimensional integrals, the noise-aided fusion rule greatly simplifies the test.

The noise-aided fusion rule is designed for the specific censoring interval of $[0, q_n]$ in the above discussion. Next, we generalize the test for the case of an arbitrary no-send region of $[t_{n1}, t_{n2}]$. It is worth mentioning that if the conditions for Theorem 5.1 or Theorem 5.2 are satisfied for X_n, these conditions are still satisfied even after adding a constant b to X_n. This can be explicitly explained from the expression for the CF of $X_n + b$:

$$\varphi_{X_n+b}(\upsilon) = e^{j\upsilon b}\varphi_{X_n}(\upsilon) \tag{5.44}$$

Similarly, an arbitrary shift of the quantization transfer characteristic in (5.10) will not affect the fulfillment of the conditions of the theorems, either. Thus, without loss of generality, we assume $t_{n1} = 0$.

The problem can be identified as the fusion of non-uniformly quantized dependent data in

which all the quantization partitions are of length q_n except for one which is $[0, t_{n2}]$. It has been demonstrated that the statistical theory of quantization can also be applied to non-uniformly quantized data [122]. We can represent the quantizer by a piecewise linear compressor (we call it a compressor, but whether it is actually a compressor or an expander depends on the ratio t_{n2}/q_n), a uniform quantizer and a piecewise linear expander which is the inverse of the compressor. Since it is the probability that data comes from a certain quantization partition that is utilized in our test, we only need to consider the piecewise linear compressor and the uniform quantizer that follows it.

A piecewise linear compressor is applied to the observations x_{nl} before quantization, whose output y_{nl} is given as

$$y_{nl} = g_n(x_{nl}) = \begin{cases} x_{nl}, & x_{nl} < 0 \\ \frac{q_n x_{nl}}{t_{n2}}, & t_{n1} \leq x_{nl} \leq t_{n2} \\ x_{nl} - t_{n2} + q_n, & x_{nl} > t_{n2} \end{cases} \tag{5.45}$$

Remark 5.3. *Since $g_n(\cdot)$ is a strictly increasing function, according to the invariance property of copulas in Theorem 1.1, the best copula that approximates the dependence among $[X_1, \ldots, X_N, X_0]$, is also the one that describes the dependence structure among $[Y_1, \ldots, Y_N, X_0]$, with the same dependence parameter.*

If we can ascertain the band-limitedness of the compressed signal Y_n, quantization theory developed for uniform quantization can be applied to the uniform quantizer which follows the compressor. Due to the piecewise linear property of the transformation in (5.45), the PDF of Y_n contains jumps at the break points of $g_n(\cdot)$. Because of such break points, the CF of Y_n may not be perfectly band-limited.

Figure 5.1 shows that the CF of raw data X_n which is Gaussian distributed and the CF of the compressed signal Y_n with different degrees of compression that is characterized by the

ratio between the length of censoring interval and the quantization step size, i.e., t_{n2}/q_n. It can be seen that Y_n is not perfectly band-limited. However, when the quantization step size is set to $q_n = 1/2$, $\phi(\upsilon) \approx 0$ for $|\upsilon| > \pi/q_n = 2\pi$, especially for small degrees of compression ($t_{n2}/q_n < 3$). The condition stated in Theorem 5.1 is satisfied very closely, when the length of the censoring interval is comparable to the quantization step size, or in other words, the break point of $g_n(\cdot)$ is not very “sharp".

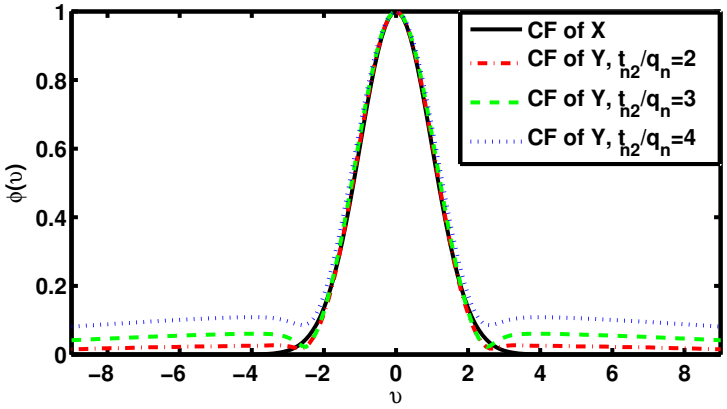

Fig. 5.1: Characteristic functions of X and Y

The compressed data, y_{nl}, is passed through a uniform quantizer with quantization step size q_n whose output is $u_{nl} = \gamma_n(y_{nl})$. This approach successfully transforms the problem of designing computationally efficient fusion rule under arbitrary communication rate constraint to the fusion of uniformly quantized data, to which we already have a noise-aided solution from the previous discussion.

Before fusion, a LPF-noise d_{nl} is added to each received data u_{nl} to create continuous observation $z_{nl} = u_{nl} + d_{nl}$ and each observation that is not received is replaced by $z_{nl} = q_n/2 + d_{nl}$. The new set of observations from sensors to be fused is $\mathbf{z}$ which are continuous. If the CF of Y_n satisfies the condition of Theorem 5.1 closely and D_n is properly designed as a LPF, we have the following relationship,

$$Z_n = Y_n + W_n + D_n \tag{5.46}$$

At the FC, the PDF of data z_{nl} under hypothesis H_i can be written as

$$f_{Z_n}(z_{nl}|H_i) = f_{Y_n}(z_{nl}|H_i) * f_{W_n}(z_{nl}) * f_{D_n}(z_{nl}) \tag{5.47}$$

The joint PDF of $\{\mathbf{z}, \mathbf{x}_0\}$ can be approximated using a copula density function as in (5.28) and the test statistic $T(\mathbf{z}, \mathbf{x}_0)$ is similar to the one in (5.43) can be derived. The fusion rule based on the addition of controlled noise greatly reduces the computational complexity, but at the expense of decreased SNR at the FC. Thus, D_n should be designed to introduce as little distortion as possible while filtering the required signal.

In may practical situations, uniform quantization is not optimal, therefore, it is important to develop models that can accommodate non-uniformly quantized-censored data. It is always possible to represent the nonuniform quantizer by combining a piecewise linear compressor, as the one in (5.45), a uniform quantizer, and a piecewise linear expander. If the CF of the compressed random variable satisfies the condition for Theorem 5.1, then the quantization theory for uniformly quantized data can be applied. An example of a floating point quantizer is given in [122], and for the number of bits used in practice, the conditions required for the quantization theorem are satisfied very closely.

5.6 Simulation Results

In this section, we provide simulation results to demonstrate the performance of our proposed fusion rules under different settings. We assume that under both hypotheses all sensors' observations are Gaussian distributed as follows

$$H_i : X_n \sim \mathcal{N}(\mu_i, \sigma^2), \quad \forall i = 0, 1 \tag{5.48}$$

for all $n = 1, \ldots, N$, where $\mathcal{N}(\mu, \sigma^2)$ denotes the univariate Gaussian distribution with mean μ and variance σ^2. It is known that the CF of a Gaussian distribution is a bell shaped function which approximates a band-limited signal. The band-limitedness will be used in the noise-aided fusion under **Scenario Q-C**. We set $[\mu_0, \mu_1] = [0, 0.5]$, $\sigma = 3$ and $L_n = -3\sigma$, $U_n = \mu_1 + 3\sigma$. We assume that under H_0, sensors' observations are independently distributed. The FC's observation is distributed according to $\mathcal{N}(0.1, 3^2)$ under hypothesis H_1 and is distributed according to $\mathcal{N}(0, 3^2)$ under H_0. Since the FC is remotely located, we assume that the FC's observation is independent of all sensors' observations under both hypotheses. Identical censoring rate constraints are applied for all sensors, i.e., $\beta_1 = \cdots = \beta_N = \beta$. Since we only focus on the fusion aspect of the detection problem for a given local censoring scheme, without loss of generality we assume the censoring region with the lower limit $t_{n1} = 0$ and t_{n2} is determined by the following censoring rate constraint

$$\int_0^{t_{n2}} f_n(x_n|H_0)dx_n = \beta.$$

We first consider a 2-sensor network, i.e., $N = 2$. Under H_1, the dependence between the two sensors' observations is generated by a Frank Copula with the corresponding Kendall's τ being 0.3. [4]

In **Scenario A-C**, sensor observations that are not in the censoring region $[0, t_{n2}]$ are transmitted to the FC. The received analog messages are directly used for deciding the true state of nature according to the copula-based GLRT in (5.16), but involving a high computational complexity. In the noise-aided GLRT, the unreceived messages are replaced by randomly generated noise at the FC before fusion. Then the detection is carried out according to (5.31). In previous research, the inter-sensor dependence is ignored in the fusion of censored data for simplicity. Under independence assumption, the test is conducted according to (5.19).

[4] Kendall's τ is a non-parametric rank-based measure of dependence, ranging from -1 to 1. Nelsen has proved the following relationship for a copula, C, and random variables $X \sim f_X(x), Y \sim f_Y(y)$ [75, p. 161]: $\tau(\boldsymbol{\phi}) = 4E[C_\phi(F_X(x), F_Y(y))] - 1$.

In **Scenario Q-C**, we set the quantization step size $q_n = \sigma/3 = 1$. Sensor observations that are not in the censoring region $[0, t_{n2}]$ are first quantized and then transmitted to the FC. After receiving the discrete messages, a copula-based GLRT is applied for deciding the true hypothesis according to (5.25). In the noise-aided fusion approach that we proposed for this scenario, our setting of the quantization step size satisfies the conditions in Theorem 5.2 closely. The LPF-noise D_n is designed to be Gaussian distributed with zero mean and variance $\sigma_D = 1$. Distortion is introduced because the CF of Gaussian distribution does not yield an ideal LPF, but is tolerable for our settings. The PDF of the signal to be fused $Z_n = Y_n + W_n + D_n$, which is nothing but the convolution of the three individual PDFs, is calculated numerically. Having the marginal PDFs, the noise-aided GLRT is conducted to test between the two hypotheses.

We set the copula library, from which the best copula is selected, to be $\mathcal{C}$= {Gaussian, Gumbel, Frank, Clayton}, which includes the true generating copula, Frank copula. The receiver operating characteristic (ROC) curves corresponding to different fusion rules for the given local censoring scheme are depicted in Figure 5.2. It can be seen that in both **Scenario A-C** and **Scenario Q-C**, our proposed noise-aided GLRTs perform comparably with copula-based GLRTs in which multidimensional integrations are evaluated numerically. Both of our proposed approaches outperform the method under Independence Assumption (IA). The detectors perform better under **Scenario A-C** compared with those under **Scenario Q-C**, which is expected since in **Scenario Q-C** by reducing data transmission through quantization, performance loss is inevitable.

To study the impact of copula misspecification on detection performance, we remove the true copula Frank copula from the copula library. The library of copulas $\mathcal{C}$, including Gaussian copula, Gumbel copula and Clayton copula, does not contain the true copula. Such a setting allows us to examine the performance of copula-based fusion under the unfavorable situation of model misspecification. The performance of our proposed fusion rules with copula misspecification is shown in Figure 5.3. The difference between the two curves corresponding

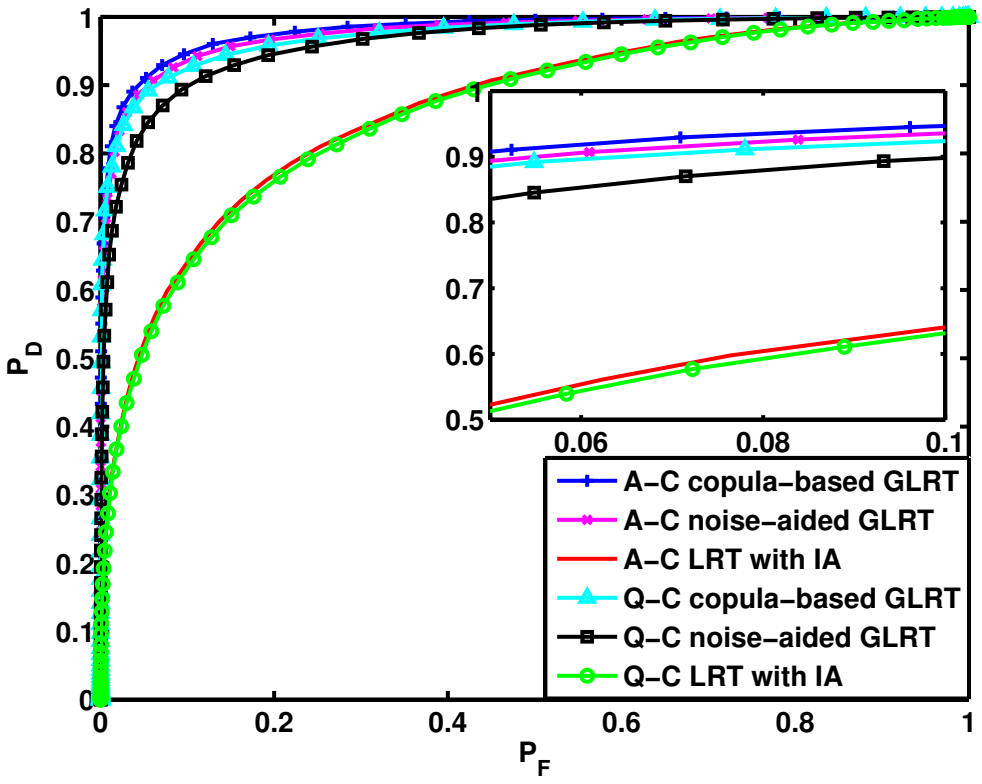

Fig. 5.2: ROCs corresponding to different fusion rules in a 2-sensor network with $\beta = 0.35$.

to the detection performance with and without copula misspecification is demonstrated to be negligible in Figure 5.3. Although the true dependence among sensor observations can be quite complex, a limited number of well defined copula families are able to characterize most dependence structures. Excluding the true copula from the copula library gives us an insight into the detection performance with misspecification in the most unfavorable situation.

The probability of correct detection P_D for a given probability of false alarm $P_F = 0.1$ is plotted as a function of censoring rate β in Figure 5.4 which captures the tradeoff between detection performance and communication efficiency in a censoring sensor network. It is observed that the performance degrades with increased censoring rate. Under **Scenario Q-C**, the gap between the performance of the noise-aided GLRT and the copula-based GLRT using brute force integration becomes larger with the increase in censoring rate. A higher censoring

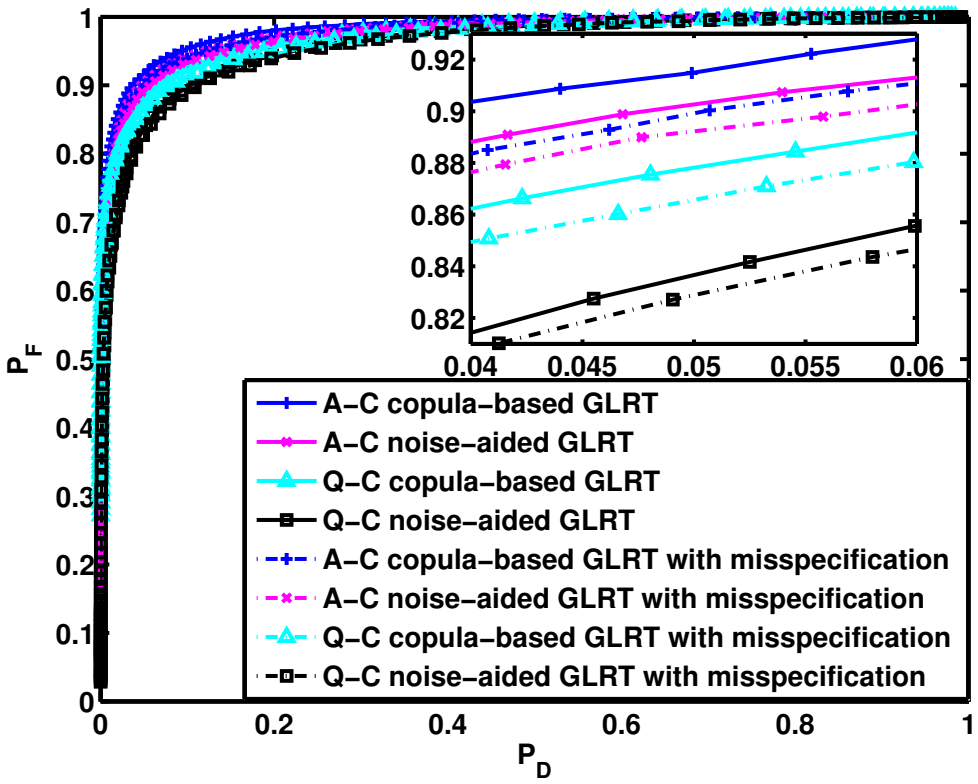

Fig. 5.3: ROCs corresponding to different copula libraries with $\beta = 0.3$: Frank copula is used to generate the data, GLRT without misspecification corresponds to the case where $\mathcal{C} = \{\text{Gaussian, Gumbel, Clayton, Frank}\}$, while GLRT with misspecification corresponds to the case where $\mathcal{C} = \{\text{Gaussian, Gumbel, Clayton}\}$.

rate leads to a higher degree of compression t_{n2}/q_n according to the piecewise-linear compressor defined in (5.45) in our noise-aided approach. Therefore, the conditions of Widrow's quantization theory are more difficult to satisfy as shown in Figure 5.1 and the main lobe of the CF of u_n can not be recovered by the process of "filtering" without noticeable distortion. Thus, attention has to be paid while applying our fusion scheme based on LPF-noise when the censoring rate constraint is high.

We also consider a multi-sensor network with $N = 3$, where dependence among sensors is

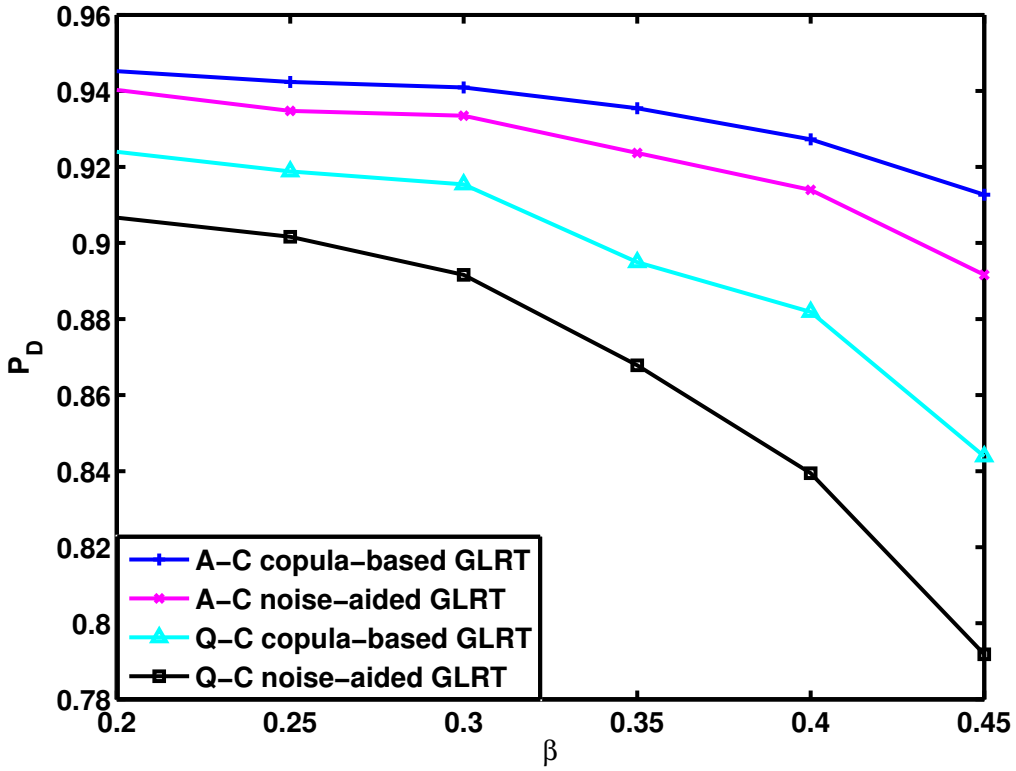

Fig. 5.4: P_D as a function of censoring rate β under **Scenario A-C** and **Scenario Q-C**

generated using a Gaussian copula with the parameter matrix given by

$$R = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$$

where $\rho = 0.25$. In this example, the copula library $\mathcal{C}$ is assumed to include Gaussian copula and t copula. The ROCs corresponding to different fusion approaches in the multi-sensor network are given in Figure 5.5. As can be seen, the performance of our proposed fusion rules that take the inter-sensor dependence into consideration is better than the test derived under independence assumption (IA). And the noise-aided GLRTs perform comparably with the copula-based GLRTs under both transmission scenarios. With the increase in the number of

sensors, computational saving of the noise-aided GLRT which transforms one N-dimensional integral to N one-dimensional integrals becomes more significant.

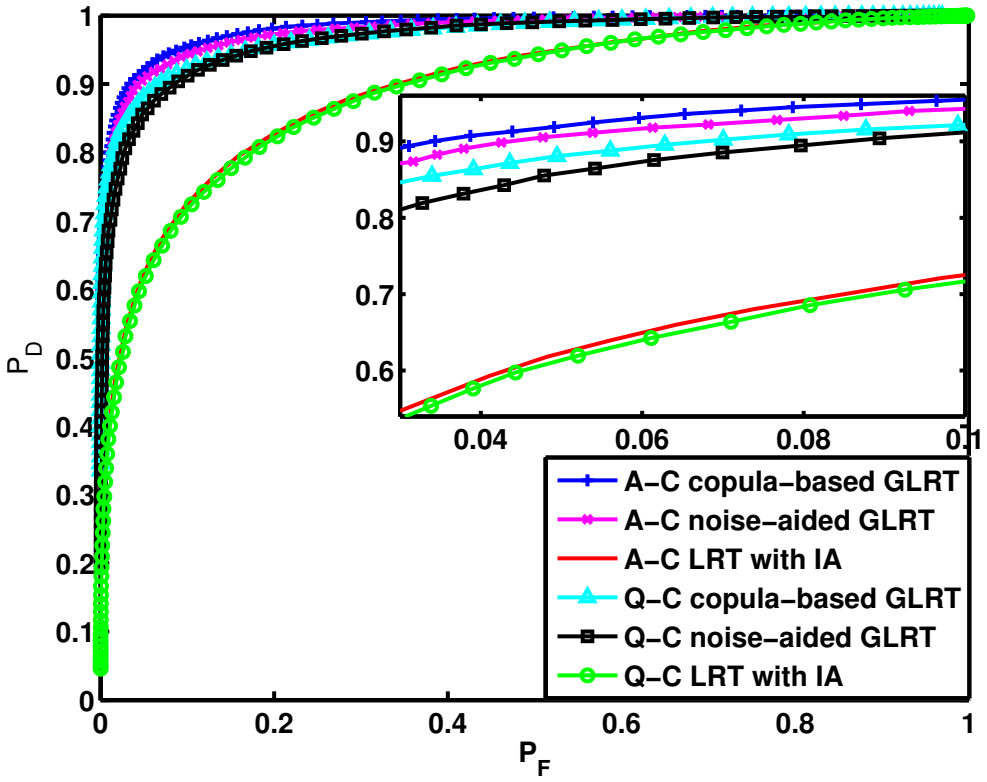

Fig. 5.5: ROCs corresponding to different fusion rules in a multi-sensor network with $\beta = 0.25$

5.7 Summary

We considered a binary hypothesis testing problem in a censoring sensor network with spatially dependent observations. Each sensor locally decides to transmit or to censor based on whether its current observation is “informative" or not. We assumed that the local censoring schemes are fixed. Two transmission scenarios were considered. In the first one, uncensored observations are transmitted directly to the FC; in the second one, uncensored observations are uniformly quantized and then transmitted. Upon the reception of messages from all transmit-

ting sensors, the FC fuses these messages with its own observations to make the final decision. The fusion rules for both analog censored data and quantized-censored data were proposed based on the characterization of unknown spatial dependence using a copula density function. The copula-based GLRT for analog censored data involves multidimensional integration, thus is expensive to compute. To address the computational issue, an alternative fusion rule that involves replacing each censored observation with an artificial noise at the FC was proposed. Another computationally efficient fusion rule by injecting controlled noise to the discrete-valued messages was presented to address a similar computational issue with the GLRT for quantized-censored data. Simulation results showed that copula-based GLRTs developed here for analog censored data and quantized-censored data and their computationally efficient versions yield significantly superior performance than the ones derived under the independence assumption.

CHAPTER 6
COALITIONAL GAMES FOR DISTRIBUTED INFERENCE IN WSNS

6.1 Motivation

In a distributed inference problem, each sensor collects observations regarding a phenomenon of interest, then shares them with other sensors or transmits them to the fusion center (FC). To reduce the energy cost for communication, the observations may be processed and compressed at the sensor before transmission. The distributed nature of wireless sensor networks induces a tradeoff between minimizing the communication cost and maintaining acceptable performance levels. Although there has been a lot of work on distributed inference, including distributed detection and distributed estimation, with conditionally independent observations, much less has been done for the case of dependent observations [22–24, 53, 54, 77, 101, 102, 123].

The spatial correlation among sensor observations is a significant characteristic which can be exploited to significantly enhance the overall network performance, including inference performance and energy efficiency. Typical applications of Wireless Sensor Networks (WSNs) require spatially dense sensor deployment in order to achieve satisfactory coverage. As a re-

sult, proximal sensors recording information about a single event are highly correlated with the degree of correlation increasing with decreasing internode separation. Such dependence among adjacent sensors or agents also exists in other intelligence aggregation networks. For example, in a crowdsourcing network, agents with the same backgrounds or having active interactions (e.g., following each other on social websites) are likely to have correlated knowledge/observations about the same event. Any network consisting of dependent agents having the ability to take measurements of the environment and making inferences based on available observations, such as wireless sensor networks, cognitive radio networks or a crowdsourcing network, falls within the scope of this work. For simplicity of presentation, we use the term "sensor" to represent an intelligent agent, which can be a real sensor, a cognitive radio, or a participating agent in a crowdsourcing network, in the remainder of this chapter. Dependence among observations may make some sensors' observations redundant. An extreme case is when two sensors' observations are completely positively correlated, one of the two sensors will become "redundant". Since transmitting "redundant" observations from battery powered sensors to remotely located FC is energy inefficient, we have an opportunity to conserve energy via local collaboration in a densely deployed sensor network.

The effect of dependent noise and hence dependent observations on Fisher Information (FI) has been studied by Yoon and Sompolinsky in [126]. The authors showed that, in the biologically relevant regime of parameters, positive correlations degrade estimation performance compared with an uncorrelated population. Sundaresan et al. [101] considered location estimation of a random signal source where they focused on improving system performance by exploiting the spatial dependence of sensor observations. Parameter estimation with dependent observations in a variety of communication scenarios was considered in [94], but was limited to the case of "geometric" dependent Gaussian noise.

Different approaches have been employed to study the detection problem with correlated observations, most of which focus on small sample size [23, 123]. It has been shown that

correlation degrades the overall performance for the case of a binary signal in equicorrelated Gaussian noise [21] as well as for the cases where correlation increases with the decrease in inter-sensor distance [20]. In parallel sensor networks, the fusion of statistically dependent observations is considered under various scenarios [53,54,102] and the design of local decision rules is investigated in [23] through the introduction of a hierarchical independence model. Performance of WSNs with correlated observations is also assessed using the theory of large deviations [22].

In this chapter, we formulate a novel distributed inference framework which exploits and utilizes the inter-sensor dependence for improved overall system performance, given the inherent tradeoff between inference performance and transmission efficiency. In such a framework, there is no FC and each individual sensor is capable of sensing and computing. Sensors form non-overlapping coalitions and collaborate by sharing their observations within a coalition. In the process of forming coalitions, each sensor selfishly aims to maximize its own inference performance, and thus the performance of the coalition to which it belongs, as will be evident later. The problem is to find a set of non-overlapping coalitions such that each sensor's inference performance is maximized under certain energy cost constraints.

In our framework, each sensor is characterized not only by its individual inference performance achieved with its own observations, but also by its dependence with other sensors in the network. To model and analyze the spatial dependence among sensor observations which might be heterogeneous, we use copula theory. Unlike the individual performance which is fixed and unchangeable no matter which coalition the sensor belongs to, its dependence with other sensors plays different roles in different coalitions. In the distributed inference problem that we are considering in this chapter, diversity gain quantifies the positive effect of dependence on inference performance, in contrast with redundancy loss, which quantifies the redundant information induced by the dependence among sensor data. Other definitions of *diversity* are available in different contexts in the signal processing literature. In cognitive radio systems,

diversity is acknowledged as the benefit of collaborative sensing and diversity order in various collaborative spectrum sensing schemes is quantitatively determined in [34]. In communication systems, *diversity* is widely adopted as an indicator of the signal-to-noise ratio (SNR) dependent behavior of inference performance based on multiple received signals [27, 110].

In this chapter, we use a game theoretical approach and formulate our collaborative distributed inference problem as a coalition formation game. Game theory has been widely applied to statistical inference, such as measurement allocation for localization [41], communication networks [43, 44, 87, 88], and spectrum sensing [86]. An iterative algorithm based on merge and split operations [9] is proposed in the literature to find a stable solution for the coalitional games discussed above.

6.1.1 Preliminary: Coalitional Game Theory

To facilitate the formulation of our problem, we introduce some basic concepts in coalitional game theory. Let $\mathcal{N} = \{1, 2, \ldots, N\}$ be a set of fixed players called the *grand coalition*. Nonempty subsets of $\mathcal{N}$ are called *coalitions*. A *collection* (in the grand coalition $\mathcal{N}$) is any family $\mathcal{S} := \{S_1, \ldots, S_m\}$ of mutually disjoint coalitions. If additionally $\cup_{j=1}^{m} S_j = \mathcal{N}$, the collection $\mathcal{S}$ is called a *partition* of $\mathcal{N}$.

Assuming a comparison relation $\triangleright$, $\mathcal{R} = \{R_1, \ldots, R_k\} \triangleright \mathcal{S} = \{S_1, \ldots, S_m\}$ means that the way $\mathcal{R}$ partitions $\mathcal{N}$, where $\mathcal{N} = \cup_{i=1}^{k} R_i = \cup_{j=1}^{m} S_j$, is preferred over the way $\mathcal{S}$ partitions $\mathcal{N}$ based on some performance measure. Pareto order can be used as a comparison relation $\triangleright$. For a collection $\mathcal{R} = \{R_1, \ldots, R_k\}$, the utility of a player j in a coalition $R_j \in \mathcal{R}$ is denoted by $\Phi_j(\mathcal{R})$, and the Pareto order is defined as follows

$$\mathcal{R} \triangleright \mathcal{S} \iff \{\Phi_j(\mathcal{R}) \geq \Phi_j(\mathcal{S}), \forall j \in \mathcal{R}, \mathcal{S}\} \tag{6.1}$$

with at least one strict inequality for a player k.

Apt and Witzel [9] proposed an abstract approach to coalition formation that focuses on simple merge-and-split rules to transform partitions of a group of players. Details of coalition formation will be introduced in Section 6.3.

6.2 System Model

We consider a physical phenomenon being continuously observed by a set of densely deployed sensors, which is represented by $\mathcal{N} = \{1, 2, \ldots, N\}$. Let θ be the parameter that denotes the phenomenon of interest in the received signal x_n for the general inference problem. When we consider a detection problem, θ represents a binary discrete variable, while in the case of parameter estimation, θ is a continuous variable that takes values in the range $[\theta_L, \theta_H]$. Each sensor's observation is represented by x_n with a known PDF $f_n(\cdot; \theta)$. We assume that the sensors in the network are heterogeneous, namely, the marginal PDFs can be distinct from each other. As has been discussed in Section 6.1, sensor observations can be highly correlated spatially and the true dependence structure can be very complex and unknown.

In a non-collaborative setting, each sensor continuously senses the environment, and locally makes inference about the unknown parameter θ solely based on its own observations. In this work, we consider a collaborative setting where collaboration exists within coalitions. Participating sensors are required to act in accordance with the following rules:

1. Sensors first form coalitions, where each sensor can only join one coalition.

2. Once the coalitions are formed, a sensor can request observations from all the other sensors in the same coalition and make an inference; it is required to transmit its observations to the other collaborating sensors upon their request.

In such a collaborative setting, each sensor, as an independent agent, aims to improve its own inference performance through collaboration with the most "useful" sensors. Let the general inference performance metric be represented by Δ. As will be shown later, in the estimation

problem Δ represents the average Fisher Information (AFI) and in the detection problem Δ represents the Kullback-Leibler Divergence (KLD). We use $\Delta_n(\mathcal{S})$ to denote the inference performance of sensor n in partition $\mathcal{S}$. The coalition formation process, namely, finding the collaboration strategy such that each selfish sensor has its performance maximized, is the focus in this chapter.

An intuitive solution would be to have all the sensors form a grand coalition such that every sensor enjoys the benefit of collaboration to the maximum extent. However, in an energy constrained network, where each sensor's energy is finite and a communication cost is incurred when it transmits observations to collaborating sensors, full collaboration is prohibitive from the energy efficiency point of view. Let r be the average number of requests initiated by each sensor in the network per unit time interval. Then, for any sensor in coalition S, the number of requests that have to be responded to within a unit time interval is $r(|S|-1)$, where $|S|$ denotes the cardinality of coalition S. We assume that energy consumption for a single transmission is E_t. The average energy consumption per unit time interval for each sensor in coalition S is $E(S) = r(|S|-1)E_t$, which increases as the coalition size increases. Let the energy consumption of a coalition be the average energy consumption per sensor in this coalition, which is the same quantity $E(S) = r(|S|-1)E_t$. Thus, from the perspective of energy efficiency, smaller coalitions are preferred. In order to guarantee adequate sensors' lifetime, we enforce the energy consumption constraint as follows

$$E(S) = r(|S|-1)E_t < \alpha, \quad \forall S \in \mathcal{S}. \tag{6.2}$$

Then the problem is to find the optimal partition $\mathcal{S}$ of the set of sensors $\mathcal{N}$ such that each sensor's inference performance is maximized subject to the energy constraint in (6.2).

$$\begin{aligned} &\max_{\mathcal{S}\in\mathcal{P}} \Delta_n(\mathcal{S}), \quad \forall n \in \mathcal{N} \\ &\text{subject to } E(S) < \alpha, \quad \forall S \in \mathcal{S} \end{aligned} \tag{6.3}$$

where $\mathcal{P}$ denotes the set of all possible partitions of $\mathcal{N}$.

For the optimization problem in (6.3), an exhaustive approach in which we search over all possible partitions will invoke a very high computational complexity. According to [90], for a network with N sensors, the total number of partitions is $O(N^N)$. Besides computational issues, a centralized exhaustive search may not be able to give us a solution to the problem in (7), since there may not exist a partition such that each sensor's performance is maximized simultaneously while the energy consumption constraint is satisfied. For the same reason, if each sensor solves its optimization problem iteratively by itself, the overall system optimization algorithm may not converge. In other centralized partition search approaches, the typical objective is to optimize the overall system performance without considering individual preferences. Since the sensors in our framework, modeled as independent agents, are selfish and act independently, they should be given the choice of making decisions on their own. This provides us the motivation to investigate distributed approaches for the sensors to make autonomous decisions, and form coalitions. To do that, we use a game theoretical approach. Before formulating the distributed inference problem as a coalition formation game, we need to define and analyze the gain and the loss of each sensor when it joins a coalition, in the context of dependent observations. The analysis is carried out respectively for the problem of estimation and detection in the following two sections.

6.3 Collaborative Distributed Estimation

In the estimation problem, the optimization problem can be formulated as the minimization of Cramer-Rao Lower Bound (CRLB), or equivalently, the maximization of Fisher Information (FI), which is given by

$$FI(\theta) = -\mathbb{E}_{\mathbf{X}}\left[\frac{\partial^2}{\partial\theta^2}\log f_{\mathbf{X}}(\mathbf{x};\theta)\right] \tag{6.4}$$

where $f_{\mathbf{X}}$ represents the joint PDF of $\mathbf{X} := [X_1, \ldots, X_N]$. Since only the range of values that θ takes is known, we assume uniform prior for the distribution of θ, which is non-informative, i.e., $f_\Theta(\theta) = \frac{1}{\theta_H - \theta_L}, \forall \theta \in [\theta_L, \theta_H]$. Thus, we define the average FI [35] as

$$I = E_\Theta \left[FI(\theta) \right] \tag{6.5}$$

For the coalition S whose set of observations is $\mathbf{x}_S := [x_n, \forall n \in S]$, the average Fisher Information (AFI) it can achieve is given as

$$I(S) = -\mathbb{E}_{\mathbf{X}_S, \Theta} \left[\frac{\partial^2 \log f_{\mathbf{X}_S}(\mathbf{x}_S; \theta)}{\partial \theta^2} \right] \tag{6.6}$$

where $f_{\mathbf{X}_S}(\cdot)$ denotes the joint distribution of $\mathbf{X}_S$ and the expectation is taken with respect to $\mathbf{X}_S$ and Θ. In the remainder of this section, the subscript of $\mathbb{E}$ is omitted for notational simplicity, unless otherwise specified.

Remark 6.1. *As an immediate result of the modus operandi of the network, the estimation performance, i.e., AFI, achievable at sensor n that is in coalition S, denoted by $I_n(S)$, equals the AFI contained in coalition S, which is denoted by $I(S)$. That is*

$$I_n(S) = I(S), \ \forall n \in S.$$

Proposition 6.1. *$I(S)$ is a nondecreasing function of the cardinality of S.*

Proof. We need to show that $I(S) \geq I(S')$, for $S' \subseteq S$. According to the definition of AFI of

coalition S in (6.6)

$$\begin{aligned} I(S) &= -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f_{\mathbf{X}_S}(\mathbf{x}_S;\theta)\right] \\ &= -\mathbb{E}_{S'}\left[\frac{\partial^2}{\partial\theta^2}\log f_{\mathbf{X}_{S'}}(\mathbf{x}_{S'};\theta)\right] + \\ &\quad \mathbb{E}_{S'}\left[-\mathbb{E}_{S\setminus S'|S'}[\frac{\partial^2}{\partial\theta^2}\log f_{\mathbf{X}_{S\setminus S'}}(\mathbf{x}_{S\setminus S'} \mid \mathbf{x}_{S'},\theta)]\right] \end{aligned} \tag{6.7}$$

where $S \setminus S^{'}$ denotes the relative complement of $S^{'}$ with respect to S, i.e., $\{n : n \in S, n \notin S^{'}\}$. It can be noted that the first term in (6.7) corresponds to the AFI of $S^{'}$, and the second term is the expected conditional AFI of $S \setminus S^{'}$. Due to the non-negativity of conditional FI, we have

$$\begin{aligned} I(S) &= I(S^{'}) + \mathbb{E}_{S'}\left[I(S \setminus S^{'}|S^{'})\right] \\ &\geq I(S^{'}) \end{aligned} \tag{6.8}$$

□

Remark 6.2. *When the transmission cost is assumed to be zero, i.e., $E_t = 0$, a grand coalition forms. It is proved in Proposition 6.1 that AFI does not decrease if more sensors join a coalition. Thus, if there is no communication cost, all the sensors will collaborate for a better estimation performance.*

It is clear from Proposition 6.1 and the definition of $E(S) = r(|S| - 1)E_t$ that, as the coalition size increases, both estimation performance in terms of $I(S)$ and the energy consumption $E(S)$ increase with it. There is a tradeoff between the estimation performance and communication efficiency. Each sensor aims to maximize its estimation performance subject to an

energy constraint. The problem is formulated as the following:

$$\begin{aligned} &\max_{\mathcal{S}\in\mathcal{P}} I_n(\mathcal{S}), \quad \forall n \in \mathcal{N} \\ &\text{s.t.} \quad r(|S|-1)E_t < \alpha, \quad \forall S \in \mathcal{S} \end{aligned} \tag{6.9}$$

where $I_n(\mathcal{S})$ represents the AFI of sensor n under partition $\mathcal{S}$. According to Remark 6.1, it is the AFI of coalition S to which sensor n belongs under $\mathcal{S}$, i.e., $I_n(\mathcal{S}) = I_n(S) = I(S)$, where $n \in S$ and $S \in \mathcal{S}$.

To analyze the effect of inter-sensor dependence on the AFI for coalition S, we express the joint PDF of observations of sensors in coalition S in terms of the marginal PDFs and copula density function c_s, as in (1.2), using copula theory. When $\log c_S(\cdot;\theta,\boldsymbol{\phi})$ is twice differentiable with respect to θ, $I(S)$ can be written as

$$\begin{aligned} I(S) &= -\mathbb{E}\left[\frac{\partial^2 \log\left(\prod_{n\in S} f_n(x_n;\theta) c_S(\cdot;\theta,\boldsymbol{\phi})\right)}{\partial\theta^2}\right] \\ &= \sum_{n\in S} I_n - \mathbb{E}\left[\frac{\partial^2 \log c_S(\cdot;\theta,\boldsymbol{\phi})}{\partial\theta^2}\right] \end{aligned} \tag{6.10}$$

where I_n represents the AFI achieved by a single sensor n in a non-collaborative setting.

Definition 6.1. *We define the second term in (6.10) as the generalized AFI (GAFI) for the copula density function* c_S [1]*, denote by* $I_c(S)$*,*

$$I_c(S) = -\mathbb{E}\left[\frac{\partial^2 \log c_S(\cdot;\theta,\boldsymbol{\phi})}{\partial\theta^2}\right]$$

$I_c(S)$ *represents the FI that is induced by the dependence structure* c_S*.*

Thus, the AFI for a coalition S can be written as the summation of AFIs of each individual sensor in S and $I_c(S)$. As to the nature of GAFI, whether it is positive (dependence contributes

[1] We call it the generalized AFI because it may not satisfy the non-negativity property of AFI.

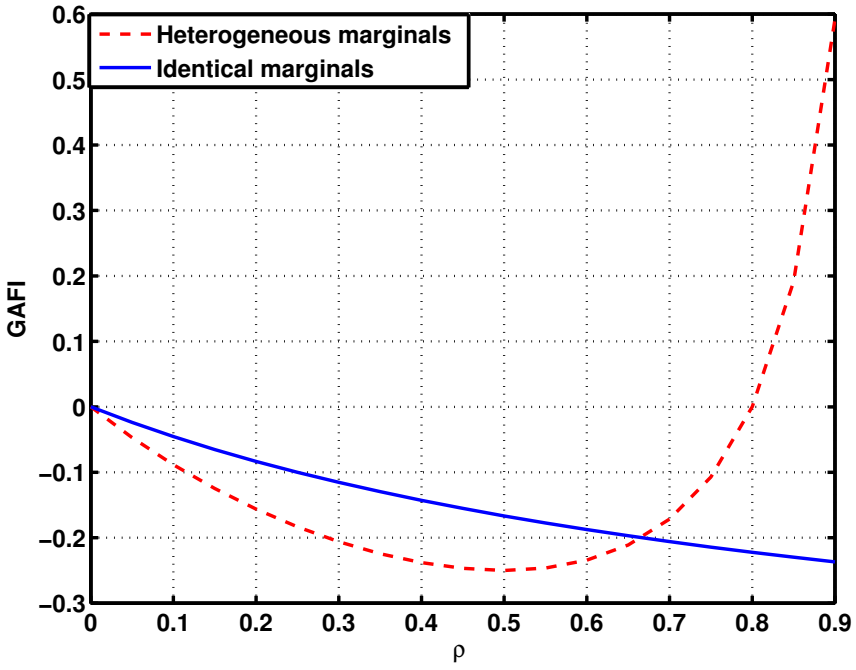

Fig. 6.1: GAFI corresponding to Gaussian copula vs. correlation coefficient ρ. The marginal distributions are Gaussian with mean $\theta_1 = \theta_2 = 1$ and variance $\sigma_1^2 = \sigma_2^2 = 4$ for the identical case; $\sigma_1^2 = 4, \sigma_2^2 = 1$ for the heterogeneous case.

to estimation performance) or negative (dependence degrades estimation performance), we present the GAFI for Gaussian copula for illustration in Figure 6.1. It is shown that for the case of identical marginal distributions, GAFI is nonpositive and monotonically decreases with an increase in the correlation coefficient ρ. More complicated behavior of GAFI, in terms of positivity/negativity and monotonicity, is observed when marginal distributions are different.

The following proposition provides some insights into the properties of GAFI for a two-sensor coalition. We assume that the dependence between the two sensors is described by a Gaussian copula and the two sensors follow univariate Gaussian distributions.

Proposition 6.2. *Let the dependence between the two random variables X and Y be described*

by a Gaussian copula [2] *with the following parameter matrix*

$$\Sigma_{XY} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

and $X \sim N(\mu_X\theta, \sigma_X)$, $Y \sim N(\mu_Y\theta, \sigma_Y)$, *where* θ *is the parameter to be estimated (Without loss of generality, let* $\left|\frac{\sigma_X}{\sigma_Y}\frac{\mu_Y}{\mu_X}\right| \leq 1$ [3]*), then we have:*

1. $I_c(X,Y)$*, the GAFI of copula* c_{XY}*, is a convex function of* ρ *and* $\min_\rho I_c(X,Y) = -\frac{\mu_Y^2}{\sigma_Y^2}$ *is reached at* $\rho = \frac{\sigma_X}{\sigma_Y}\frac{\mu_Y}{\mu_X}$*;*

2. $I_c(X,Y) \leq 0$ *when* ρ *lies between* 0 *and* $\frac{2\mu_X\mu_Y\sigma_X\sigma_Y}{\mu_X^2\sigma_Y^2+\mu_Y^2\sigma_X^2}$.

3. *When* $\frac{\sigma_X}{\sigma_Y}\frac{\mu_Y}{\mu_X} = 1$, $I_c(X,Y) \geq 0$ *for* $\rho \in [-1,0]$ *and* $I_c(X,Y) < 0$ *for* $\rho \in (0,1]$. *Furthermore,* $I_c(X,Y)$ *is a monotonically decreasing function of* ρ*.*

Proof. According to the definition of GAFI in (6.10)

$$\begin{aligned} I_c(X,Y) &= -\mathbb{E}\left[\frac{\partial^2 \log c_{XY}(F_X(x;\theta), F_Y(y;\theta);\rho)}{\partial\theta^2}\right] \\ &= \frac{-1}{\sigma_X^2\sigma_Y^2(1-\rho^2)}\{2\rho\mu_X\mu_Y\sigma_X\sigma_Y \\ &\quad -\rho^2(\mu_X\sigma_Y^2 + \mu_Y\sigma_X^2)\} \end{aligned} \tag{6.11}$$

From the second derivative of $I_c(X,Y)$, it can be seen that [4]

$$\frac{\partial^2 I_c(X,Y)}{\partial\rho^2} \geq 0, \quad \forall\rho \in (-1,1)$$

[2] The Gaussian copula is defined as

$$C_{\text{Gaussian}}(\mathbf{u}|\Sigma) = \Phi_\Sigma(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_N)),$$

where, Φ_Σ denotes the multivariate normal CDF and Φ denotes the univariate normal CDF.

[3] The inequality $\left|\frac{\sigma_X}{\sigma_Y}\frac{\mu_Y}{\mu_X}\right| \leq 1$ implies that the individual AFI of Y is less than or equal to that of X, i.e., $I_Y \leq I_X$, since $I_Y = \frac{\mu_Y^2}{\sigma_Y^2}$ and $I_X = \frac{\mu_X^2}{\sigma_X^2}$

[4] The dependence of $I_c(X,Y)$ on correlation coefficient ρ is not made explicit for notational convenience.

thus, $I_c(X,Y)$ is convex. By setting

$$\frac{\partial I_c(X,Y)}{\partial \rho} = 0$$

and knowing that

$$\left| \frac{\sigma_X}{\sigma_Y} \frac{\mu_Y}{\mu_X} \right| \leq 1$$

we get

$$\rho^* = \frac{\sigma_X}{\sigma_Y} \frac{\mu_Y}{\mu_X}, \quad I_c(X,Y)^* = -\frac{\mu_Y}{\sigma_Y^2}$$

Thus, the minimum value of $I_c(X,Y)$ is $I_c(X,Y)^* = -\frac{\mu_Y^2}{\sigma_Y^2}$ and it is obtained at $\rho^* = \frac{\sigma_X}{\sigma_Y} \frac{\mu_Y}{\mu_X}$.

By setting (6.11) equal to zero, we get two solutions:

$$\rho_1 = 0, \quad \rho_2 = \frac{2\mu_X \mu_Y \sigma_X \sigma_Y}{\mu_X^2 \sigma_Y^2 + \mu_Y^2 \sigma_X^2}$$

Combined with the convexity of the function, it can be concluded that $I_c(X,Y) \leq 0$ when $\rho \in [\rho_2, 0]$ if $\rho_2 < 0$ and $\rho \in [0, \rho_2]$ if $\rho_2 \geq 0$.

By letting $\sigma_X \mu_Y = \sigma_Y \mu_X$ in (6.11), the conclusions in 3) can be directly derived. □

Remark 6.3. *When $\rho = 0$, then $I_c(X,Y) = 0$, meaning that there is no impact of dependence on estimation performance and the AFI of the coalition is solely the summation of individual AFIs of X and Y, i.e., $I_X + I_Y$.*

Remark 6.4. *Assuming that X has a larger individual AFI compared with Y, i.e., $I_X > I_Y$, when $\rho = \frac{\sigma_X}{\sigma_Y} \frac{\mu_Y}{\mu_X}$, $I_c(X,Y)$ equals $-I_Y$. In this case, X gains nothing in estimation performance by collaboration.*

Remark 6.5. *A sensor n prefers to collaborate with sensor m with which it has a positive $I_c(X_n, X_m)$ than sensor k with which it has a negative $I_c(X_n, X_k)$, when sensor m and sensor k have identical individual AFI. For sensor n, sensor m is more "favorable" than sensor k*

in the sense that the inter-sensor dependence with m brings "diverse" information while that with k brings "redundant" information.

Definition 6.2. *We define $-I_c(X,Y)\mathbb{1}_{I_c(X,Y)<0}$ to be pairwise redundancy loss denoted as $I_{rl}(X,Y)$, and $I_c(X,Y)\mathbb{1}_{I_c(X,Y)\geq 0}$ to be pairwise diversity gain denoted as $I_{dg}(X,Y)$, where $\mathbb{1}_{\{\cdot\}}$ is an indicator function.*

The definitions of diversity gain and redundancy loss allow for a better characterization of the different roles that pairwise inter-sensor dependence may play. The GAFI of multivariate copulas can be analyzed using *vines* which is a graphical method of constructing multivariate copulas [65, 97]. In this formulation, we essentially establish a hierarchical, pairwise dependence relation, which can be expressed through copulas. The joint PDF of N random variables expressed in terms of a D-vine decomposition is given by:

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{n=1}^{N} f(x_n) \prod_{j=1}^{N-1} \prod_{k=1}^{N-j} c_{j,j+k|\overline{j}}(F(x_j|\,\mathbf{x}_{\overline{j}}), F(x_{j+k}|\,\mathbf{x}_{\overline{j}})) \tag{6.12}$$

where $\overline{j} = [j+1, \ldots, j+k-1]$, $\mathbf{x}_{\overline{j}} = [x_{j+1}, \ldots, x_{j+k-1}]$. $c_{j,j+k|\overline{j}}(\cdot)$ is basically the pairwise copula between j and $j+k$ conditioned on $\overline{j}$. Thus, a multivariate copula is decomposed into the product of bivariate conditional copulas. The corresponding GAFI $I_c(S)$ of the copula in any coalition S can be written as:

$$\begin{aligned} I_c(S) &= \sum_{j=1}^{|S|-1} \sum_{k=1}^{|S|-j} I_c(X_j, X_{j+k}|\,\mathbf{X}_{\overline{j}}) \\ &= \sum_{j=1}^{|S|-1} \sum_{k=1}^{|S|-j} I_{dg}(X_j, X_{j+k}|\,\mathbf{X}_{\overline{j}}) \\ &\quad - \sum_{j=1}^{|S|-1} \sum_{k=1}^{|S|-j} I_{rl}(X_j, X_{j+k}|\,\mathbf{X}_{\overline{j}}) \\ &= I_{dg}(S) - I_{rl}(S) \end{aligned} \tag{6.13}$$

where $I_{dg}(S)$ and $I_{rl}(S)$ respectively represent the diversity gain and redundancy loss in the coalition S. Each of them is a summation of pairwise diversity gains or pairwise redundancy losses in coalition S. Until now, we have quantified the benefit and cost of forming a coalition S incurred by dependent sensor observations in the problem of distributed estimation. In the next section, the counterparts of diversity gain and redundancy loss for the distributed detection problem will be investigated.

6.4 Collaborative Distributed Detection

In the detection problem, θ is a bi-valued variable which takes the value θ_0 under hypothesis H_0 and takes the value θ_1 under hypothesis H_1. in this chapter, we employ Kullback-Leibler Divergence (KLD) as the performance metric. KLD can be interpreted as the error exponent in the Neyman-Pearson framework, which means that the probability of missed detection goes to zero exponentially with the number of observations at a rate equal to KLD. Thus, KLD characterizes the asymptotic detection performance. We denote KLD by D and define it as follows

$$D = \mathbb{E}_{H_0}\left[\log \frac{f_{\mathbf{X}}(\mathbf{x}\,|H_0)}{f_{\mathbf{X}}(\mathbf{x}\,|H_1)}\right] \tag{6.14}$$

where $\mathbb{E}_{H_0}[\cdot]$ denotes the expectation taken with respect to the joint distribution of $\mathbf{X}$ under hypothesis H_0. For a coalition S, the detection performance, in terms of KLD, is $D(S)$.

$$D(S) = \mathbb{E}_{X_S|H_0}\left[\log \frac{f_{\mathbf{X}_S}(\mathbf{x}_S\,|H_0)}{f_{\mathbf{X}_S}(\mathbf{x}_S\,|H_1)}\right] \tag{6.15}$$

Remark 6.6. *The KLD for sensor n, i.e., $D_n(S)$, in a coalition S is the same for all $n \in S$.*

Similar to Remark 6.1 for the estimation problem, we can write

$$D_n(S) = D(S),\ \forall n \in S$$

Proposition 6.3. *D(S) is nondecreasing in* $|S|$.

Proof. In order to prove that $D(S)$ does not decrease if new members join the existing coalition, we need to show that for any $S^{'} \subseteq S$, $D(S^{'}) \leq D(S)$.

$$\begin{aligned}
D(S) &= \int \log \frac{f_{\mathbf{x}_S}(\mathbf{x}_S\,|H_0)}{f_{\mathbf{x}_S}(\mathbf{x}_S\,|H_1)} f_{\mathbf{x}_S}(\mathbf{x}_S\,|H_0)\mathrm{d}\,\mathbf{x}_S \\
&= \int \log \frac{f_{\mathbf{x}_{S'}}(\mathbf{x}_{S'}\,|H_0)}{f_{\mathbf{x}_{S'}}(\mathbf{x}_{S'}\,|H_1)} f_{\mathbf{x}_{S'}}(\mathbf{x}_{S'}\,|H_0)\mathrm{d}\,\mathbf{x}_{S'} \\
&+ \int \log \frac{f_{\mathbf{x}_{S\setminus S'}}(\mathbf{x}_{S\setminus S'}\,|\,\mathbf{x}_{S'}, H_0)}{f_{\mathbf{x}_{S\setminus S'}}(\mathbf{x}_{S\setminus S'}\,|\,\mathbf{x}_{S'}, H_1)} f_{\mathbf{x}_S}(\mathbf{x}_S\,|H_0)\mathrm{d}\,\mathbf{x}_S \\
&= D(S^{'}) + \mathbb{E}_{\mathbf{x}_{S'}\,|H_0}\left[D(S\setminus S^{'})\right] \\
&\geq D(S^{'}) \qquad (6.16)
\end{aligned}$$

The last inequality is because of the non-negativity property of conditional KLD. □

Remark 6.7. *A grand coalition forms when communication cost is zero, i.e.,* $E_t = 0$.

It is noted that, as $|S|$ increases, both $D(S)$ and $E(S)$ increase, indicating a tradeoff between the detection performance and energy consumption. In our formulation, each sensor selfishly aims to maximize its own detection performance, i.e., the KLD when using shared observations within the coalition to which it belongs, subject to an energy constraint. The problem can be formulated as the following

$$\begin{aligned}
&\max_{\mathcal{S}\in\mathcal{P}} D_n(\mathcal{S}), \quad \forall n \in \mathcal{N} \\
&\text{s.t.} \quad r(|S|-1)E_t < \alpha, \quad \forall S \in \mathcal{S} \qquad (6.17)
\end{aligned}$$

where $D_n(\mathcal{S})$ represents the KLD of sensor n under partition $\mathcal{S}$. According to Remark 6.6, it is the KLD of coalition S to which sensor n belongs under $\mathcal{S}$, i.e., $D_n(\mathcal{S}) = D_n(S) = D(S)$, where $n \in S$ and $S \in \mathcal{S}$.

The effect of inter-sensor dependence on the KLD can be analyzed by expressing the joint PDF of the observations of sensors in coalition S in terms of the marginal PDFs and copula density function c_s. By copula theory, the KLD corresponding to $\mathbf{X}_S$ can be written as

$$\begin{aligned} D(S) &= \int \log \frac{\prod_{n\in S} f_n(x_n|H_0) c_S(\cdot|\boldsymbol{\phi}_0, H_0)}{\prod_{n\in S} f_n(x_n|H_1) c_S(\cdot|\boldsymbol{\phi}_1, H_1)} f_{\mathbf{X}_S}(\mathbf{x}_S\,|H_0)\mathrm{d}\,\mathbf{x}_S \\ &= \sum_{n\in S} D_n + \mathbb{E}_{\mathbf{X}_S\,|H_0}\left[\log \frac{c_S\left(F_n(x_n|H_0), n\in S|\boldsymbol{\phi}_0, H_0\right)}{c_S\left(F_n(x_n|H_1), n\in S|\boldsymbol{\phi}_1, H_1\right)}\right] \end{aligned} \tag{6.18}$$

where D_n is the KLD achieved by sensor n with its own observations in a non-collaborative setting and $\boldsymbol{\phi}_i$ is the dependence parameter of the copula density under hypothesis H_i, $i = 0, 1$. The second term in (6.18) is due to inter-sensor dependence and if we assume independent sensors, i.e., $c_S(\cdot) = 1$, then the second term becomes zero.

Definition 6.3. *We define the second term in (6.18) as the Generalized KLD (GKLD), denoted by $D_c(S)$, which measures the KL distance between the two joint distributions introduced by the dependence structure.*

$$D_c(S) = \mathbb{E}_{\mathbf{X}_S\,|H_0}\left[\log \frac{c_S\left(F_n(x_n|H_0), n\in S|\boldsymbol{\phi}_0, H_0\right)}{c_S\left(F_n(x_n|H_1), n\in S|\boldsymbol{\phi}_1, H_1\right)}\right] \tag{6.19}$$

Thus, the KLD between the two joint distributions of sensor observations in S under hypotheses H_0 and H_1 can be decomposed into two terms, as shown in (6.18). The first term represents the summation of KLDs corresponding to individual sensors in S and the second term is $D_c(S)$. We call $D_c(S)$ the Generalized KLD, because the arguments of $c_S(\cdot|H_0)$ and $c_S(\cdot|H_1)$ are different and thus violate the standard definition of KLD. In Figure 6.2, the GKLDs corresponding to different copulas are plotted against Kendall's τ. For each curve, there exists

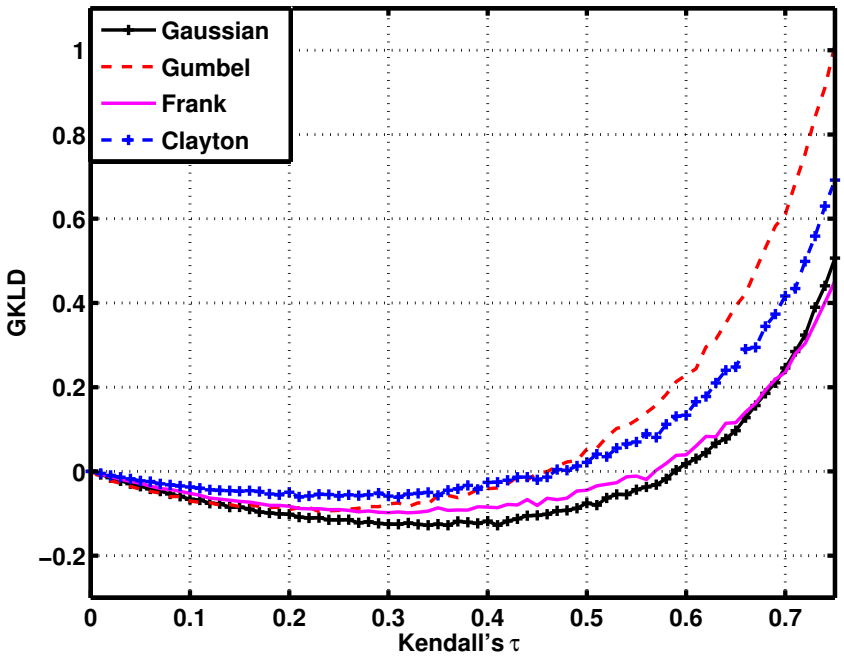

Fig. 6.2: GKLD corresponding to different copulas vs. Kendall's τ. Gaussian marginals are assumed. Means are assumed to be $\theta_1 = \theta_2 = 2$ under H_0 and $\theta_1 = \theta_2 = 1$ under H_1. Variances are assumed to be $\sigma_1^2 = 4, \sigma_2^2 = 1$ under both hypotheses.

a single τ^* that divides $\tau \in [0, 1]$ into two intervals, each corresponding to positive or negative GKLD respectively. Since Kendall's τ is only a scalar summarization of the "amount" of dependence, the behavior of GKLD varies for different copula models (dependence structures).

The following proposition provides insights into GKLD corresponding to a Gaussian copula in a coalition consisting of two sensors that follow univariate Gaussian distributions.

Proposition 6.4. *Consider the following the detection problem*

$$H_1 : X \sim N(\theta_1, \sigma_X), Y \sim N(\theta_1, \sigma_Y), c_{Gaussian}(\Sigma_{XY})$$

$$H_0 : X \sim N(\theta_0, \sigma_X), Y \sim N(\theta_0, \sigma_Y), c_{Gaussian}(\Sigma_{XY})$$

where

$$\Sigma_{XY} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

and $\theta_1 \neq \theta_0$*. Without loss of generality, let* $\sigma_X \leq \sigma_Y$ [5]*, then we have:*

1. $D_c(X,Y)$*, the GKLD corresponding to the Gaussian copula* c_{XY}*, is a convex function of* ρ *and* $\min_\rho D_c(X,Y) = -\frac{(\theta_1-\theta_0)^2}{2\sigma_Y^2}$ *is reached at* $\rho = \frac{\sigma_X}{\sigma_Y}$*;*

2. $D_c(X,Y) \leq 0$ *for* ρ *between* 0 *and* $\frac{2\sigma_X\sigma_Y}{\sigma_Y^2+\sigma_X^2}$*.*

3. *For* $\sigma_X = \sigma_Y$*,* $D_c(X,Y) \geq 0$ *for* $\rho \in [-1,0]$ *and* $D_c(X,Y) < 0$ *for* $\rho \in (0,1]$ *and it is a monotone decreasing function of* ρ*,.*

Proof. According to the definition of GKLD, we have

$$\begin{aligned} & D_c(X,Y) \\ & \quad = \mathbb{E}_{XY|H_0}\left[\log \frac{c\left(F_X(x|H_0), F_Y(y|H_0)|\Sigma_{XY}\right)}{c\left(F_X(x|H_1), F_Y(y|H_1)|\Sigma_{XY}\right)}\right] \\ & \quad = \frac{(\theta_1-\theta_0)^2}{2\sigma_X^2\sigma_Y^2(1-\rho^2)}\left[\rho^2(\sigma_Y^2+\sigma_X^2) - 2\rho\sigma_X\sigma_Y\right] \end{aligned} \tag{6.20}$$

where $c(\cdot|\Sigma_{XY})$ represents the Gaussian copula parameterized by $\boldsymbol{\phi} = \Sigma_{XY}$. It can be shown that

$$\frac{\partial^2 D_c(X,Y)}{\partial \rho^2} \geq 0, \quad \forall \rho \in (-1,1)$$

Thus, the convexity is proved. By setting

$$\frac{\partial D_c(X,Y)}{\partial \rho} = 0$$

[5] $\sigma_X \leq \sigma_Y$ implies that $D_X \geq D_Y$, since $D_X = \frac{(\theta_1-\theta_0)^2}{2\sigma_X^2}$ and $D_Y = \frac{(\theta_1-\theta_0)^2}{2\sigma_Y^2}$

and knowing that $\sigma_X \leq \sigma_Y$, we get

$$\rho^* = \frac{\sigma_X}{\sigma_Y}, \quad D_c(X,Y)^* = -\frac{(\theta_1 - \theta_0)^2}{2\sigma_Y^2}$$

which when combined with the convexity of the function yields that $D_c(X,Y)^*$ is the minimum point.

By setting (6.20) equal to zero, we get two solutions:

$$\rho_1 = 0, \quad \rho_2 = \frac{2\sigma_X\sigma_Y}{\sigma_Y^2 + \sigma_X^2}$$

Due to the convexity of the function, it can be concluded that $D_c(X,Y) \leq 0$ when $\rho \in [\rho_1, \rho_2]$.

When $\sigma_X = \sigma_Y = \sigma$,

$$D_c(X,Y) = -\frac{(\theta_1 - \theta_0)^2}{\sigma^2}\frac{\rho}{1+\rho}$$

It can be shown that $\frac{\partial D_c(X,Y)}{\partial \rho} \leq 0, \forall \rho \in (-1,1)$ and the sign of $D_c(X,Y)$ is the same as that of $-\rho$. □

Remark 6.8. *When X and Y are independently distributed, i.e., $\rho = 0$, then $D_c(X,Y) = 0$, meaning that KLD is solely the summation of individual KLDs of X and Y.*

Remark 6.9. *Assuming that X has larger individual KLD compared with Y, i.e., $D_X > D_Y$ when $\rho = \frac{\sigma_X}{\sigma_Y}$, $D_c(X,Y)$ equals $-D_Y$. In this case, X does not improve its detection performance by forming a coalition with Y.*

Remark 6.10. *A sensor n would prefer to collaborate with sensor m with which it has a positive $D_c(X_n, X_m)$ than sensor k with which it has a negative $D_c(X_n, X_k)$, when sensor m and k have identical individual performance, i.e., $D_m = D_k$. To sensor n, sensor m is more "favorable" in the sense that the dependence between sensor n and m results in a larger total KLD, and thus a better asymptotic detection performance.*

Definition 6.4. *We define* $-D_c(X,Y)\mathbb{1}_{D_c(X,Y)<0}$ *to be pairwise redundancy loss of GKLD, denoted as* $D_{rl}(X,Y)$*, and define* $D_c(X,Y)\mathbb{1}_{D_c(X,Y)\geq 0}$ *to be pairwise diversity gain of GKLD denoted as* $D_{dg}(X,Y)$*.*

Although the expressions of the pairwise redundancy loss and diversity gain depend on the specific problem that we are considering, these definitions capture the intrinsic characteristics of a sensor network with dependent observations and quantify the impact of the dependence in collaboration.

According to (6.12), a multivariate copula can be decomposed into the product of bivariate conditional copulas. Therefore, $D_c(S)$, the GKLD introduced by the copula in any coalition S can be written as:

$$\begin{aligned} D_c(S) &= \sum_{j=1}^{|S|-1}\sum_{k=1}^{|S|-j} D_c(X_j, X_{j+k}|\,\mathbf{X}_{\bar{j}}) \\ &= \sum_{j=1}^{|S|-1}\sum_{k=1}^{|S|-j} D_{dg}(X_j, X_{j+k}|\,\mathbf{X}_{\bar{j}}) \\ &- \sum_{j=1}^{|S|-1}\sum_{k=1}^{|S|-j} D_{rl}(X_j, X_{j+k}|\,\mathbf{X}_{\bar{j}}) \\ &= D_{dg}(S) - D_{rl}(S) \end{aligned} \tag{6.21}$$

$D_{dg}(S)$ represents the diversity gain in the coalition S and $D_{rl}(S)$ represents the amount of redundant information included in coalition S. By noting that $D_{dg}(S)$ and $D_{rl}(S)$ are nonnegative and nondecreasing function of $|S|$, we can view $D_{dg}(S)$ together with $\sum_{n\in S} D_n$ as the gain of forming S, while $D_{rl}(S)$ as the cost, along with the communication cost $E(S)$. In the following section, a coalition formation game for distributed inference is formulated based on the quantification of dependence-based diversity gain and redundancy loss.

6.5 Game Formulation and Properties

We propose a coalitional game defined by the pair $(\mathcal{N}, V)$ to model our collaborative inference problem, where $\mathcal{N}$ is the set of players (all sensors) and V is a mapping such that for every coalition S, $V(S)$ is a closed convex subset of $\mathbb{R}^S$ that contains the payoffs that players in S receive. The value of any coalition S can be mapped to a set V of payoff vectors.

Before defining $V(S)$, we should first define the value of a coalition, denoted by $v(S)$, which is supposed to characterize the tradeoff between the gain from forming coalition S and the corresponding loss. As has been discussed in the previous two sections, the gain of forming coalition S comes from the summation of individual sensor inference performance and the dependence related diversity gain, while the overall cost of forming coalition S is due to the dependence related redundancy loss and the communication cost. In order to present a generalized game theoretical approach to the distributed inference problem, we use a unified notation Δ to represent the AFI in the estimation problem, i.e., I and the KLD in the detection problem, i.e., D. We define the value of a coalition $v(S)$, as an increasing function of the gain of forming coalition S, i.e., $\sum_{n\in S}\Delta_n + \Delta_{dg}(S)$, and a decreasing function of the costs $\Delta_{rl}(S)$, and $C(S)$ which is a function of the communication cost $E(S)$:

$$v(S) = \left[\sum_{n\in S}\Delta_n + \Delta_{dg}(S)\right] - [\Delta_{rl}(S) + C(S)] \tag{6.22}$$

where $C(S)$, as a function of $E(S)$, depends on the communication constraint α in the original optimization problem given in (7). It captures the tradeoff between inference performance and energy consumption, since the first three terms in (26) correspond to the inference performance of S and the last term corresponds to the communication cost. There are certain properties that a well designed cost function $C(S)$ should satisfy, here we use the logarithmic barrier penalty

function given in [17]

$$C(S) = \begin{cases} -1/t \cdot \log(1 - \frac{E(S)}{\alpha}) & \text{if } E(S) < \alpha \\ +\infty & \text{otherwise} \end{cases} \tag{6.23}$$

where α is the constraint on $E(S)$, and t is a control parameter. The above cost function is an increasing function of $E(S)$ for $E(S) < \alpha$, while it goes to infinity when $E(S) \geq \alpha$. Through the cost function in (6.23), the constraint that $E(S) < \alpha$ in (6.3) is enforced, since for the coalitions that do not satisfy this constraint, the utility $v(S)$ is $-\infty$.

Proposition 6.5. *The payoff for each sensor in coalition S is equal to the utility of the coalition, i.e., $\Phi_n(S) = v(S), \forall n \in S$, where $\Phi_n(S)$ denotes the payoff of sensor n in the coalition S.*

Proof. The value of a coalition S defined in (6.22) is a function of its inference performance and its average energy consumption $E(S)$. According to Remarks 6.1 and 6.6, the AFI or KLD for every sensor in S is given by the AFI and KLD of the coalition. And it is known that transmission cost $E(S)$ of every sensor in S is the average transmission cost of the coalition. Hence, the coalition value $v(S)$ is also the payoff of each player in it. □

Now, we have a coalitional game $(\mathcal{N}, V)$, where $V(S)$ is a singleton set (hence closed and convex)

$$V(S) := \{\mathbf{\Phi}(S) | \Phi_n(S) = v(S), \forall n \in S\} \tag{6.24}$$

Remark 6.11. *The proposed coalitional game $(\mathcal{N}, V)$ for the distributed inference problem is a nontransferable utility game.*

A distributed algorithm for the above coalition formation game among sensors is described next.

6.5.1 Coalition Formation Algorithm

For autonomous coalition formation, we propose a distributed algorithm based on two simple rules called *merge* and *split* [9] that allow us to modify a partition $\mathcal{S}$ of the set $\mathcal{N}$.

Merge Rule: Merge any set of coalitions $\{S_1, \ldots, S_m\}$, where $\{\cup_{j=1}^{m} S_j\} \triangleright \{S_1, \ldots, S_m\}$, therefore, $\{S_1, \ldots, S_m\} \rightarrow \{\cup_{j=1}^{m} S_j\}$.

Split Rule: Split any coalition $\{\cup_{j=1}^{m} S_j\}$, where $\{S_1, \ldots, S_m\} \triangleright \{\cup_{j=1}^{m} S_j\}$, thus $\{\cup_{j=1}^{m} S_j\} \rightarrow \{S_1, \ldots, S_m\}$.

Remark 6.12. *Every iteration of the merge and split rules terminates.*

Let us assume that the dependence information is known at the local sensors, and they autonomously form coalitions through merge and split operations. Let the initial partition be $\mathcal{S} = \{S_1, \ldots, S_m\}$.

repeat

$\mathcal{R} =$ Merge($\mathcal{S}$): coalitions in $\mathcal{S}$ merge according to the merge rule, until no further merge occurs

$\mathcal{S} =$ Split($\mathcal{R}$): coalitions in $\mathcal{R}$ split according to the split rule, until no further split occurs.

until No merge or split occurs

Merge operations are first applied. Given an initial partition $\mathcal{S} = \{S_1, \ldots, S_m\}$, suppose S_1 seeks to collaborate with S_2. If the condition for merge is satisfied, a new coalition $S_1 := S_1 \cup S_2$ is formed, otherwise, $S_1 := S_1$ and S_1 attempts to merge with another coalition that shares a mutual benefit in merging. The algorithm is repeated for the remaining S_i until all the coalitions have made their merge decisions. The resulting partition $\mathcal{R}$ is then subject to a split process in a similar way. Then, successive merge-and-split processes go on until the iterations terminate.

The stability of this resulting network structure can be investigated using the concept of a defection function $\mathbb{D}$ [9, 86].

Definition 6.5. *A defection function $\mathbb{D}$ is a function which is associated with each partition $\mathcal{T}$. A partition $\mathcal{T} = \{T_1, \ldots, T_m\}$ is $\mathbb{D}$-stable if no group of players is interested in leaving $\mathcal{T}$ when the players who leave can only form the coalitions allowed by $\mathbb{D}(\mathcal{T})$.*

A partition $\mathcal{T} = \{T_1, \ldots, T_m\}$ of $\mathcal{N}$ is $\mathbb{D}_{hp}$-stable, if no players in $\mathcal{T}$ are interested in leaving $\mathcal{T}$ through *merge-and-split* to form other partitions in $\mathcal{N}$. A partition $\mathcal{T}$ is $\mathbb{D}_c$-stable, if no players in $\mathcal{T}$ are interested in leaving $\mathcal{T}$ through *any operation* to form other coalitions in $\mathcal{N}$ [86].

$\mathbb{D}_{hp}$-stability can be thought of as a state of equilibrium where no coalitions have an incentive to pursue coalition formation through merge or split. The following remark has been shown in [8].

Remark 6.13. *A partition is $\mathbb{D}_{hp}$-stable if and only if it is the outcome of iterating the merge-and-split rules.*

Remark 6.14. *For the proposed $(\mathcal{N}, V)$ collaborative distributed inference game, the proposed merge-and-split algorithm converges to a $\mathbb{D}_{hp}$-stable partition.*

It is known that if $\mathcal{T}$ is $\mathbb{D}_c$-stable, then $\mathcal{T}$ is the outcome of every iteration of the merge-and-split rules and it is a unique $\mathbb{D}_c$-stable partition [9]. Nonetheless, a $\mathbb{D}_c$-stable partition does not always exist. The $\mathbb{D}_c$-stable partition $\mathcal{S} = \{S_1, \ldots, S_m\}$ of the whole space $\mathcal{N}$ exists if and only if the following two conditions hold:

1. For each $i \in \{1, \ldots, m\}$ and each pair of disjoint coalitions A and B, such that $\{A \cup B\} \subseteq S_i$, we have $\{A \cup B\} \triangleright \{A, B\}$.

2. For the partition $\mathcal{S} = \{S_1, \ldots, S_m\}$, a coalition $G \subset \mathcal{N}$ formed of players belonging to different $S_i \in \mathcal{S}$ is $\mathcal{S}$-incomplete if for no $i \in \{1, \ldots, m\}$ we have $\{G\}[\mathcal{S}] \triangleright \{\mathcal{G}\}$, where $\{G\}[\mathcal{S}] = \{\mathcal{G} \cap \mathcal{S}_i, \forall i \in \{1, \ldots, m\}\}$ is the projection of coalition, $\mathcal{G}$ in partition $\mathcal{S}$.

For the proposed $(\mathcal{N}, v)$ collaborative distributed inference game, the proposed merge-and-split algorithm converges to a $\mathbb{D}_{hp}$-stable partition. If a $\mathbb{D}_c$-stable partition exists, then

the $\mathbb{D}_{hp}$-stable partition is also the optimal $\mathbb{D}_c$-stable partition. The existence of the $\mathbb{D}_c$-stable partition cannot always be guaranteed, although the algorithm would always converge to this partition when it exists.

Remark 6.15. *For the proposed $(\mathcal{N}, V)$ collaborative distributed inference game, the proposed merge-and-split algorithm converges to the optimal $\mathbb{D}_c$-stable partition, if such a partition exists. Otherwise, the proposed algorithm converges to a $\mathbb{D}_{hp}$-stable partition.*

6.6 Simulation Results

In this section, we present some simulation results to examine the performance of our proposed game theoretical approach to the collaborative distributed inference problem. We consider a wireless sensor network with N sensors deployed in a $[0, 1.5] \times [0, 1.5]$ square area of interest. Let the location of sensor n be denoted by $\mathbf{s_n} = [s_{n1}, s_{n2}]$. The amount of dependence measured in terms of Kendall's τ between any two sensors n and m follows the power exponential model [115]

$$\tau(d_{n,m}) = \mathrm{e}^{-d_{n,m}^2}, \tag{6.25}$$

where $d_{n,m} = \|\mathbf{s}_n - \mathbf{s}_m\|$ is the distance between nodes n and m respectively located at coordinates $\mathbf{s}_n$ and $\mathbf{s}_m$.

We first consider an 8-sensor network where each sensor's observation follows Gaussian distribution with mean θ and variance σ_n^2. The inter-sensor dependence is described by a Gaussian copula with correlation matrix Σ whose elements are obtained from the pairwise τ in (6.25). Let $\mathbf{s_S}$ denote the location of the signal source which is $[0.75, 0.75]$ in this experiment. The variance of each sensor's observation is inversely proportional to the distance between the sensor and the signal source, i.e., $\sigma_n^2 = 1/|\mathbf{s_n} - \mathbf{s_S}|$. We set $rE_t = 1$ and $\alpha = 4$, thus, according to (6.2), the largest coalition size that satisfies the energy efficiency constraint is $|S| = 4$.

In the problem of estimation, the unknown parameter θ is assumed to be within $[-1, 1]$. The AFI of coalition S is given as

$$I(S) = \mathbf{1}^T \Big(\operatorname{diag}(\boldsymbol{\sigma}_S) \Sigma_S \operatorname{diag}(\boldsymbol{\sigma}_S) \Big)^{-1} \mathbf{1} \tag{6.26}$$

where $\boldsymbol{\sigma}_S$ is a vector consisting of standard deviations of all sensors in S, i.e., $\boldsymbol{\sigma}_S = [\sigma_n, \forall n \in S]$; $\mathbf{1}$ is an all one vector with dimension $|S|$ by 1, and Σ_S is the correlation matrix of coalition S.

In the detection problem, we set the parameters under hypothesis H_0 and H_1 to be $\theta_0 = 0$ and $\theta_1 = \sqrt{2}$. The KLD corresponding to a coalition S is

$$D(S) = \mathbf{1}^T \Big(\operatorname{diag}(\boldsymbol{\sigma}_S) \Sigma_S \operatorname{diag}(\boldsymbol{\sigma}_S) \Big)^{-1} \mathbf{1} \tag{6.27}$$

With the above setting, the AFI and KLD have exactly the same expression. Thus, we present the simulation results without distinguishing between the problems of estimation and detection.

In the initialization step, each sensor is set to be a coalition by itself, i.e.,

$$\mathcal{S} = \big\{ \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \{8\} \big\}.$$

By applying the proposed merge-and-split algorithm iteratively, sensors' payoffs increase with each merge/split operation until no further merge/split happens as shown in Figure 6.3. Eventually three coalitions $\{\{1, 2, 4, 7\}, \{3, 5\}, \{6, 8\}\}$ are formed and within each coalition, sensors have the same payoff. It can be seen in Figure 6.4 that each coalition contains sensors that are physically apart, and thus statistically less dependent so that redundancy loss is avoided and diversity gain is introduced to the largest degree. Also, the sensors closer to the signal source, who already have a good individual performance, form smaller coalitions, while the distantly located sensors form relatively larger coalitions to improve their performance. A $\mathbb{D}_c$-stable

solution is not guaranteed in this example, since the resulting partitions are different with different initializations as shown in Figure 6.5. With each iteration of merge-and-split, the overall payoff [6] of sensors increases, until no further merge or split occurs as shown in Figure 6.5.

We further consider a heterogeneous sensor network consisting of 28 sensors deployed in the same $[0, 1.5] \times [0, 1.5]$ square area of interest. We assume that there are 14 sensors whose observations follow Gaussian distribution with mean θ and unit variance, while the other 14 sensors' observations follow exponential distribution parameterized by θ. Within each Monte Carlo trial, the sensor locations are generated independently according to uniform distribution, through which the correlation matrix is obtained according to the dependence model in (6.25). A student's t copula parameterized by the correlation matrix with the degree of freedom $\nu = 4$ is used to generate the dependence among sensors. In this experiment, we assume that the true dependence among heterogeneous sensors is unknown and we will demonstrate how copula

[6] We use "overall payoff" to imply the payoff averaged over all sensors. The term "overall" will continually be used with the same implication in the later part of this section.

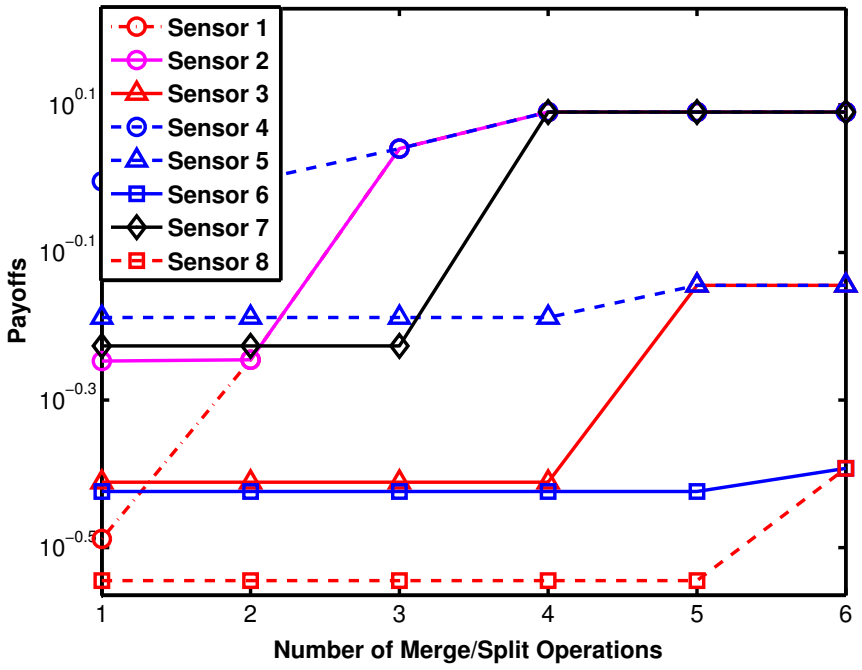

Fig. 6.3: Individual sensor payoffs vs. number of merge/split operations.

theory enables the characterization of the dependence structure in a heterogeneous network, upon which the distributed coalition formation approach is built.

Before the coalition formation process, we first learn the dependence among sensors by collecting $L = 1000$ training data from each sensor, denoted by $\mathbf{x} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ where $\mathbf{x}_n = [x_{n1}, \ldots, x_{nL}]$. We apply a semi-parametric approach in which the marginal CDFs are obtained non-parametrically (due to unknown marginal parameters) and the dependence parameter corresponding to a certain copula is estimated parametrically. The uniform random variables (marginal CDFs) in the copula density are evaluated using empirical probability integral transform (EPIT) as follows

$$\hat{F}_n(\cdot) = \frac{1}{L}\sum_{l=1}^{L} \mathbb{1}_{x_{nl}<\{\cdot\}}$$
$$\hat{u}_{nl} = \hat{F}_n(x_{nl}) \tag{6.28}$$

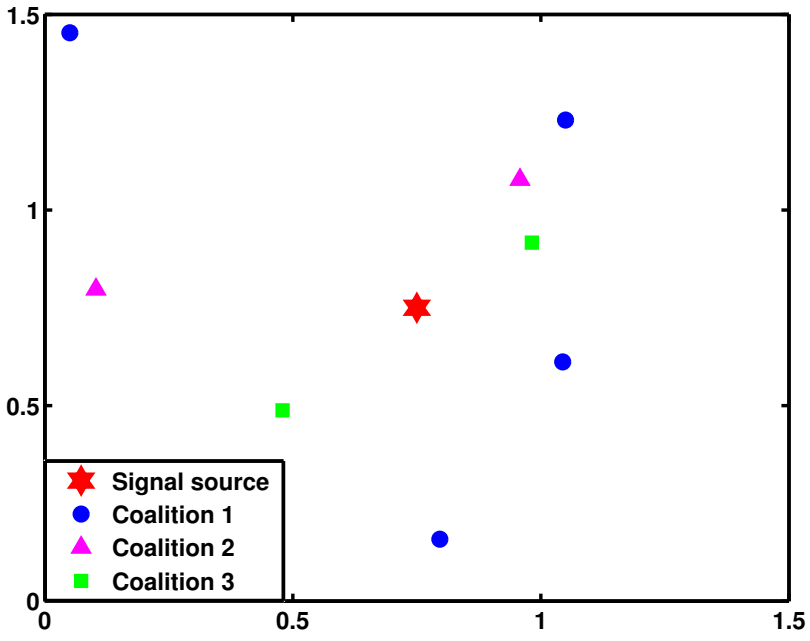

Fig. 6.4: The deployment of the 8-sensor network and the final partition.

For a given copula c, its corresponding parameter is obtained by Maximum Likelihood (ML) estimation as follows

$$\hat{\boldsymbol{\phi}} = \arg\max_{\boldsymbol{\phi}} \sum_{l=1}^{L} \log c(\hat{u}_{1l}, \ldots, \hat{u}_{Nl}|\boldsymbol{\phi}) \tag{6.29}$$

Besides estimating the copula parameter, selecting the best copula from a predefined finite set of copulas, $\mathcal{C}$, is also one important step. We use a minimum description length (MDL) [47] based approach for model selection. MDL techniques of model selection are based on the principle that the model that achieves the best compression is the model best suited, from the available alternatives, to describe the data. There are several criteria available under the MDL framework such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Stochastic Information Complexity (SIC) and Normalized Maximum Likelihood (NML). We use BIC in our experiment and the predefined copula library is set to be

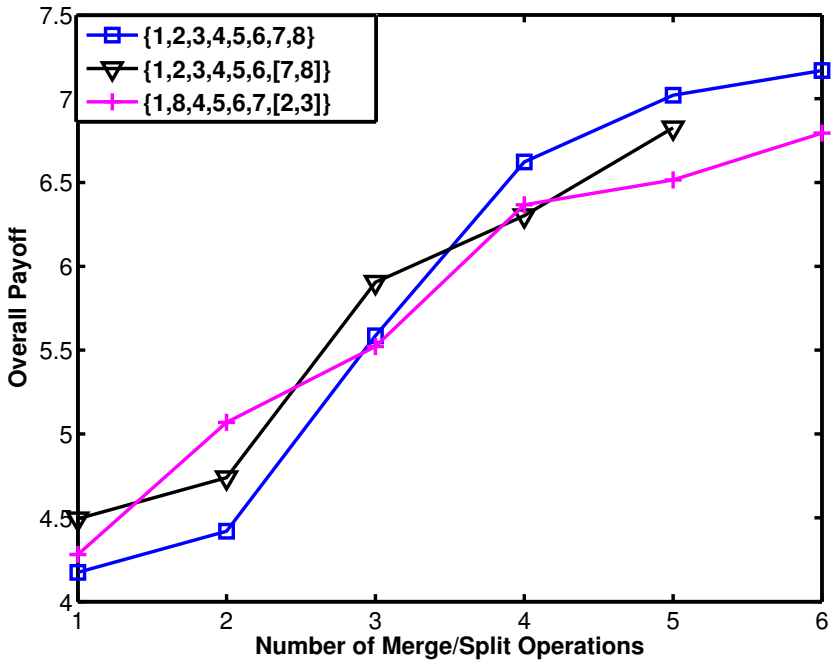

Fig. 6.5: The overall payoff vs. number of merge/split operations for different initialization of the merge-and-split algorithm.

$\mathcal{C} = \{\text{Gaussian}, \text{Student's t}\}$. The best copula c^* is selected as follows

$$c^* = \arg\max_{c \in \mathcal{C}} \sum_{l=1}^{L} \log c(\hat{u}_{1l}, \ldots, \hat{u}_{Nl}|\hat{\boldsymbol{\phi}}) + \frac{d}{2} \log L \tag{6.30}$$

where d is the dimension of $\boldsymbol{\phi}$, and $\hat{\boldsymbol{\phi}}$ is obtained through (6.29). A more detailed discussion of copula selection is available in [52, 98].

The performance corresponding to different coalition formation approaches for this particular sensor deployment is evaluated. A total of 100 Monte Carlo trials are conducted and the performance is averaged over these trials. Within each trial, the dependence structure is learnt using our proposed copula-based approach for the generated sensor deployment. We compare our coalition formation approach, in which the dependence structure of the network is first modeled using a multivariate copula function and then the coalition formation game is formed and solved, with the approach under Independence Assumption (IA). In the coalition formation under IA, the unknown dependence among heterogeneous sensors is ignored, thus for any coalition S, both the dependence-related diversity gain and redundancy loss are zero. Thus, in the utility function in (6.22) there are only two terms left, the first term concerning individual sensor performance and the last term concerning communication cost. The same merge-and-split algorithm is then applied to solve the game under IA. We also compare our proposed distributed algorithm based on coalition formation game with the approach of random coalition formation. In the random coalition formation method, a partition is randomly selected from the set of all partitions that satisfy the communication constraint with equality [7].

In the estimation problem, θ is assumed to be within $[-1, 1]$. Figure 6.6 shows the overall estimation performance of our proposed distributed coalition formation approach, compared with the random coalition formation approach and coalition formation under IA. As the con-

[7] The equality is to ensure a maximized inference performance, since the inference performance is nondecreasing in coalition size, according to Proposition 6.1 and Proposition 6.3. We make the coalition size to be exactly α, except for the one that may include less than α due to the fact that the total number of sensors N may not be an integer multiple of α.

straint on communication cost gets looser (α increases), the overall estimation performance becomes better for all methods. However, since our approach explores and utilizes inter-sensor dependence during the coalition formation process, it achieves much better performance than the other two methods. The overall communication costs, defined as $1/|\mathcal{S}| \sum_{S \in \mathcal{S}} E(S)$, are plotted in Figure 6.7, which demonstrates the superiority of our method in terms of communication efficiency. It has to be noted that, the average communication cost corresponding to our distributed coalition formation method is not the maximum communication cost that is allowed by the predefined constraint. It reflects the true cost of communication of the resulting partition, which may be much less than the maximum cost allowed. The curves for the coalition formation under IA and random coalition formation are the same in Figure 6.7, since in both approaches each coalition ends up with the maximum number of sensors allowed by the constraint, if possible.

In the detection problem, we set $\theta_0 = 1$ and $\theta_1 = 2.4$. Superior overall detection perfor-

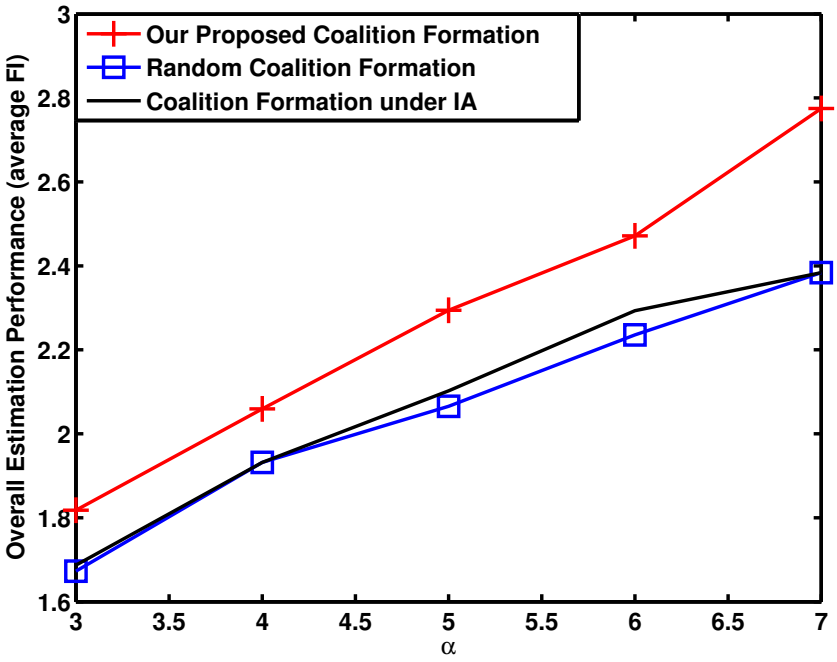

Fig. 6.6: Overall estimation performance vs. communication constraint α.

mance of the partitions resulting from our proposed coalition formation approach is shown in Figure 6.8, in comparison with the coalition formation under IA and random coalition formation. It is noted that for the other two approaches, the overall KLDs drop when α goes from 6 to 7. This is because with increasing coalition size the amount of redundant information increases due to inter-sensor dependence, especially when the redundancy loss is not taken into consideration in the coalition formation process as in random and IA coalition formation processes. The overall actual communication costs versus α are plotted in Figure 6.9, demonstrating a better communication efficiency of our approach. In our distributed algorithm, when two coalitions are unable to contribute much to each other in inference performance due to their dependence (or redundancy loss incurred), they will not merge into a new coalition. Thus, it forces the coalition to seek cooperation with other coalitions to which it can contribute more, or where it is highly valued due to the diverse information that it is able to bring in. In this

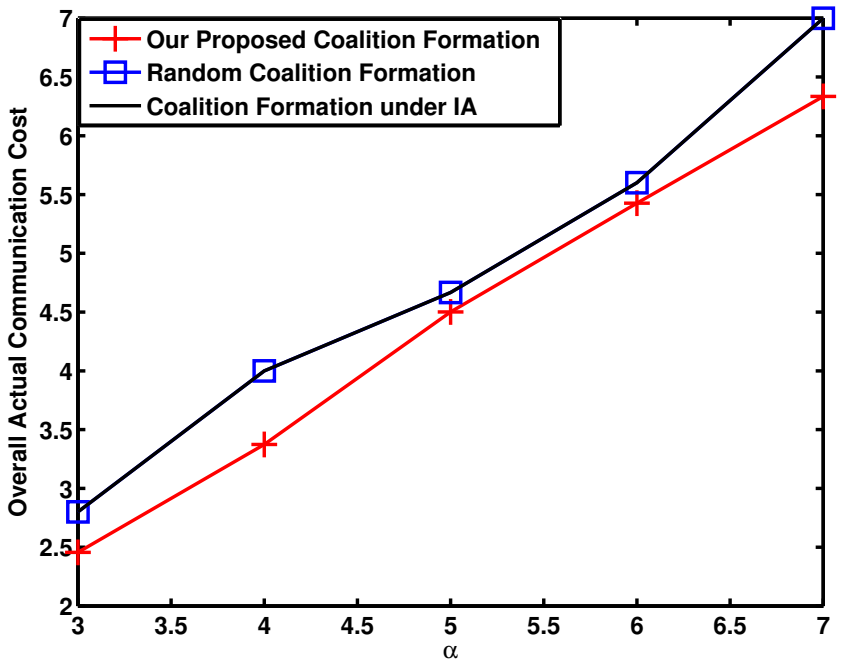

Fig. 6.7: Overall actual communication cost vs. communication constraint α in estimation problem. (The curves for Coalition Formation under IA and Random Coalition Formation are overlapped.)

way, the overall diversity gain is increased while redundancy loss is decreased. By formulating the distributed inference problem as a coalition formation game and solving the game using an iterative algorithm, we are able to obtain better system performance in terms of both inference performance and energy efficiency, compared with the coalition formation under IA and the random coalition formation scheme.

In numerous practical scenarios, sensor networks are subject to changes. For example, sensors embedded in people's cellphones change locations frequently. New sensors joining or existing sensors quitting also contributes to the time varying nature of the network. The distributed nature of our proposed coalition formation method in which sensors form coalitions automatically, makes it suitable for mobile networks with time-varying configurations.

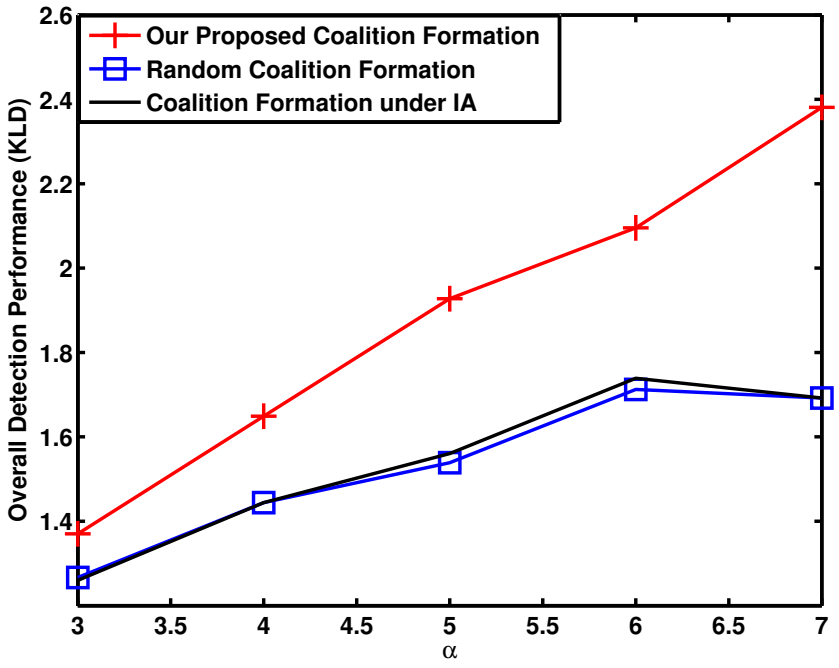

Fig. 6.8: Overall detection performance vs. communication constraint α.

6.7 Summary

In this chapter, we investigated a collaborative distributed inference problem in an energy constrained wireless sensor network with dependent observations. In the collaborative setting, sensors form coalitions and share observations within the coalition for improved inference performance. We focused on the formation of non-overlapping collaborating coalitions such that each sensor's performance is maximized while the energy constraint is satisfied. To analyze the benefit and cost of forming a certain coalition, we used copula theory to describe the dependence among observations, which provided "redundancy" and "diversity" aspects of inter-sensor dependence, respectively for the problem of estimation and detection. We defined GAFI and GKLD to quantify the diversity gain and redundancy loss in forming a coalition due to inter-sensor dependence. A coalition formation game was proposed for the distributed infer-

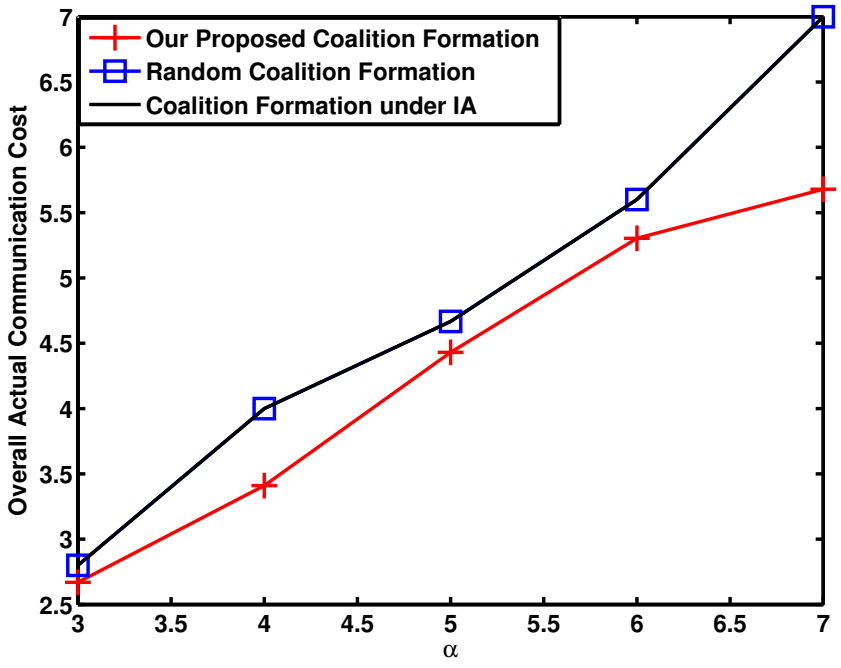

Fig. 6.9: Overall actual communication cost vs. communication constraint α in detection problem. (The curves for Coalition Formation under IA and Random Coalition Formation are overlapped.)

ence problem. A merge-and-split algorithm was utilized for our coalitional game and the stability of the outcome of our proposed algorithm was analyzed. Finally, numerical results were provided to demonstrate the performance of our game theoretical approach. Further investigation of the dependence-related concepts of diversity gain and redundancy loss in inference problems under different scenarios is to be conducted in the future work.

CHAPTER 7
SUMMARY AND FUTURE DIRECTIONS

7.1 Summary

In this thesis, we investigated several inference problems in networks with heterogeneous information sources which are statistically dependent. Both centralized and decentralized inference problems were considered. Due to the statistical dependence among those heterogeneous information sources, we proposed several methods that take such dependence into consideration to improve the inference performance.

The problem of centralized detection with heterogeneous sensor data was investigated in the Neyman-Pearson framework. We applied copula theory to characterize the dependence among sensor observations. According to the true nature of the dependence among sensor measurements, we considered two kinds of situations. When the dependence structure is non-stationary, meaning that the copula which describes the dependence structure may change over time, we proposed a sample-wise copula selection rule which selects the copula with the maximum likelihood for each sample. The optimality of this method over existing approaches was proved theoretically and demonstrated through simulations. When the true dependence follows a mixture model, meaning that each observation is generated according to a certain copula with

corresponding probability, we designed a detector based on a mixture of copulas. Expectation Maximization (EM) algorithm was applied to iteratively find the estimates of unknown parameters in the mixture model. Theoretical proofs and simulation results were also provided to demonstrate the performance of our proposed approaches.

Inspired by the exponentially growing amount of social media data, we considered the inference with both sensor data and social media data. To jointly characterize the sensor data and social media data, we proposed a copula-based model for multivariate time series. We used an ARMA-GARCH model to capture the temporal dynamics of each time series and then used copula theory to build the joint distribution among their standardized residual terms. The proposed model is very flexible in capturing a wide range of temporal and spatial dependence structures. It provides a powerful tool for characterizing heterogeneous time series (sensor data and social data) by accommodating very complex dependence structures. By applying such a model to jointly characterize the GT returns and stock returns of Apple Inc., we were able to have a better prediction of the stock price. Its application in other inference tasks assisted by social media data, such as flu prediction, was also shown.

In distributed networks, where only local decisions are transmitted to the FC for global decision making, we investigated the classification of random signals. With the introduction of a “hidden" random variable, we derived the necessary conditions for the optimal sensor compression rules and fusion rule under conditionally dependent observations. An iterative algorithm, which is easy to implement, was proposed to generate the sensor rules and the fusion rule. Convergence and asymptotic optimality of the proposed algorithm were also proved. We applied the method that we proposed in this work to a modulation classification problem and it was shown to exhibit better performance than the other approaches, as shown in the numerical results.

A binary hypothesis testing problem was considered in a censoring sensor network with spatially dependent observations. In this formulation, each sensor decides to transmit or to

censor based on whether its current observation is “informative" or not. Two transmission scenarios were considered. In the first one, uncensored observations are transmitted directly to the FC; while in the other, uncensored observations are uniformly quantized and then transmitted. The fusion rules for both analog censored data and quantized-censored data were proposed based on the characterization of unknown spatial dependence using a copula density function. The copula-based GLRT for analog censored data involves multidimensional integrations, thus is expensive to compute. To address the computational issue, an alternative fusion rule that involves replacing each censored observation with an artificial noise at the FC was proposed. Another computationally efficient fusion rule by injecting controlled noise to the discrete-valued messages was presented to address a similar computational issue with copula-based GLRT for quantized-censored data. Simulation results showed that copula-based GLRTs developed here for analog censored data and quantized-censored data and their computationally efficient versions yield significantly superior performance than the ones derived under the independence assumption.

Finally, we investigated a collaborative distributed inference problem in an energy constrained wireless sensor network with dependent observations. In this collaborative setting, sensors form non-overlapping coalitions and share observations within the coalition for improved inference performance. We focused on the collaboration strategy such that each sensor’s performance is maximized while the energy constraint is satisfied. To analyze the benefit and cost of forming a certain coalition, we used copula theory to describe the dependence among observations, which provided “redundancy” and “diversity” aspects of inter-sensor dependence. We defined GAFI and GKLD to quantify the diversity gain and redundancy loss in forming a coalition due to inter-sensor dependence. A coalition formation game was proposed for the distributed inference problem. A merge-and-split algorithm was utilized for our coalitional game and the stability of the outcome of our proposed algorithm was analyzed. Numerical results were provided to demonstrate the performance of our game theoretical ap-

proach.

7.2 Future Directions

Some promising directions for future work are listed in the following:

1. In Chapter 2, we have provided the sample-wise copula selection approach for the case where the parameters of copula functions are estimated, without theoretical justification. Although it is an intuitive approach inspired by the solution under the case of known parameters, a theoretical justification, under some mild conditions, is desired. As a future work, one can investigate the conditions under which the sample-wise copula selection approach with estimated copula parameters outperforms a single copula approach with estimated parameters.

2. In this thesis, we have investigated the problem of copula selection and parameter estimation in different scenarios. However, in both centralized and decentralized frameworks, we focused on the problem of binary hypotheses testing. It is of interest to study the problem of copula selection and parameter estimation for multiple hypotheses testing problem under dependent observations. Copula selection approaches for the specific task of multiple hypotheses testing, as the copula selection method that is embedded in the GLRT for the detection problems, is worth pursuing as a future work.

3. We have studied dependence modeling among different time series and how to make a better prediction by taking such intermodal dependence into consideration in Chapter 3. It is of great interest to study the causality of these time series, and figure out which is the cause time series and which is the effect time series. Granger's definition of causality has been widely applied due to its simplicity and robustness. However, due to the heterogeneity of the time series (from sensors and social media), the linearity of VAR (vector autoregressive)-based Granger causality based test can be easily violated. Since we have

been applying copula to describe nonlinear dependence in the previous research, we propose, as a future work, the development of Copula-Granger algorithm to conduct the causality analysis on multiple heterogeneous time series concerning a certain unfolded event of interest.

4. In Chapter 5, we designed the fusion rules for detection with analog and discrete data in a censoring sensor network for given censoring schemes at the local sensors. The designed fusion algorithms take inter-sensor dependence into consideration to improve detection performance. It is of interest to further investigate optimal or near optimal local censoring schemes and study how inter-sensor dependence affects the design of the censoring rules. Due to the complex coupling between sensor rules introduced by dependence, optimal censoring scheme for each individual sensors may not be available. One can investigate necessary conditions for optimal censoring rules and design iterative algorithms to search for the rules according to the necessary conditions, in the future. The hierarchical structure for dependent sensor observations which was introduced in Chapter 4 can be applied to facilitate the design of local censoring rules. But, the major challenge in the censoring problem comes from the fact that unlike the fusion rule for the distributed AMC problem which is only a function of local decisions, the fusion rule of censored data is not only a function of local transmissions but also of local censoring rules.

5. In Chapter 6, the dependence-related definitions of diversity gain and redundancy loss were explored using copula theory where only spatial dependence was considered, while temporal independence was assumed. One can, in the future, explore the diversity gain and redundancy loss corresponding to temporal dependence. Especially in a sequential detection problem, temporal dependence affects the average sensing time that is required to reach a detection decision and the different roles it plays need to be investigated. The copula-based model for multiple dependent time series has been proposed in Chapter 3,

and it can be applied for facilitating further investigation of diversity and redundancy in inference problems with both temporal and spatial dependence.

REFERENCES

[1] V. Aalo and R. Viswanathan, "On distributed detection with correlated sensors: two examples," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 25, no. 3, pp. 414–421, 1989.

[2] ——, "On distributed detection with correlated sensors: two examples," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 25, no. 3, pp. 414–421, 1989.

[3] A. S. Abu-Romeh and D. L. Jones, "Decentralized detection in censoring sensor network under correlated observations," *EURASIP Journal on Advances in Signal Processing*, 2010.

[4] P. Addesso, S. Marano, and V. Matta, "Sequential sampling in sensor networks for detection with censoring nodes," *Signal Processing, IEEE Transactions on*, vol. 55, no. 11, pp. 5497–5505, Nov 2007.

[5] M. Anderson, T. Adali, and X.-L. Li, "Joint Blind Source Separation with Multivariate Gaussian Model: Algorithms and Performance Analysis," *Signal Processing, IEEE Transactions on*, vol. 60, no. 4, pp. 1672–1683, 2011.

[6] S. Appadwedula, V. Veeravalli, and D. Jones, "Energy-efficient detection in sensor networks," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 4, pp. 693–702, 2005.

[7] ——, "Decentralized detection with censoring sensors," *Signal Processing, IEEE Transactions on*, vol. 56, no. 4, pp. 1362–1373, 2008.

[8] K. R. Apt and T. Radzik, "Stable Partitions in Coalitional Games," *CoRR*, vol. abs/cs/0605132, 2006.

[9] K. R. Apt and A. Witzel, "A Generic Approach to Coalition Formation," *International Game Theory Review*, vol. 11, no. 3, pp. 347–367, 2009.

[10] S. Asur and B. A. Huberman, "Predicting the future with social media," in *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, vol. 1, 2010, pp. 492–499.

[11] M. J. Beal, N. Jojic, and H. Attias, "A graphical model for audiovisual object tracking," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 7, pp. 828–836, 2003.

[12] B. Bhanu and V. Govindaraju, Eds., *Multibiometrics for Human Identification*. New York, NY: Cambridge University Press, 2011.

[13] J. A. Bilmes *et al.*, "A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," *International Computer Science Institute*, vol. 4, no. 510, p. 126, 1998.

[14] S. Blackman and R. Popoli, *Design and analysis of modern tracking systems*. Norwood, MA: Artech House, 1999.

[15] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.

[16] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[18] T. Butz and J.-P. Thiran, "From error probability to information theoretic (multi-modal) signal processing," *Signal Processing*, vol. 85, no. 5, pp. 875–902, May 2005.

[19] M. Çetin, L. Chen, I. Fisher, J. W., A. T. Ihler, R. L. Moses, M. J. Wainwright, and A. S. Willsky, "Distributed fusion in sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 42–55, 2006.

[20] J. Chamberland and V. Veeravalli, "Wireless sensors in distributed detection applications," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 16–25, May 2007.

[21] J.-F. Chamberland and V. Veeravalli, "Decentralized detection in sensor networks," *Signal Processing, IEEE Transactions on*, vol. 51, no. 2, pp. 407–416, Feb 2003.

[22] ——, "How dense should a sensor network be for detection with correlated observations?" *Information Theory, IEEE Transactions on*, vol. 52, no. 11, pp. 5099–5106, Nov 2006.

[23] H. Chen, B. Chen, and P. Varshney, "A new framework for distributed detection with conditionally dependent observations," *Signal Processing, IEEE Transactions on*, vol. 60, no. 3, pp. 1409–1419, 2012.

[24] Q. Cheng, P. Varshney, J. Michels, and C. Belcastro, "Distributed fault detection with correlated decision fusion," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 45, no. 4, pp. 1448–1465, Oct 2009.

[25] U. Cherubini, E. Luciano, and W. Vecchiato, *Copula Methods in Finance*. Prentice Hall, 2004.

[26] N. M. Correa, T. Adali, Y.-O. Li, and V. D. Calhoun, "Canonical Correlation Analysis for Data Fusion and Group Inferences: Examining applications of medical imaging data." *IEEE Signal Process Mag*, vol. 27, no. 4, pp. 39–50, June 2010.

[27] S. Cui, J.-J. Xiao, A. Goldsmith, Z.-Q. Luo, and H. Poor, "Estimation diversity and energy efficiency in distributed sensing," *Signal Processing, IEEE Transactions on*, vol. 55, no. 9, pp. 4683–4695, Sept 2007.

[28] T. Damarla, "Hidden markov model as a framework for situational awareness," in *Proc. 11th International Conference on Information Fusion*, 2008, pp. 1–7.

[29] M. Davy and A. Doucet, "Copulas: a new insight into positive time-frequency distributions," *IEEE Signal Processing Letters*, vol. 10, no. 7, pp. 215–218, Jul. 2003.

[30] A. D'Costa, V. Ramachandran, and A. Sayeed, "Distributed classification of Gaussian space-time sources in wireless sensor networks," *Selected Areas in Communications, IEEE Journal on*, vol. 22, no. 6, pp. 1026–1036, August 2004.

[31] O. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," *Communications, IET*, vol. 1, no. 2, pp. 137–156, April 2007.

[32] O. Dobre and F. Hameed, "Likelihood-based algorithms for linear digital modulation classification in fading channels," in *Electrical and Computer Engineering, 2006. CCECE '06. Canadian Conference on*, May 2006, pp. 1347–1350.

[33] E. Drakopoulos and C.-C. Lee, "Optimum multisensor fusion of correlated local decisions," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 27, no. 4, pp. 593–606, 1991.

[34] D. Duan, L. Yang, and J. Principe, "Cooperative diversity of spectrum sensing for cognitive radio systems," *Signal Processing, IEEE Transactions on*, vol. 58, no. 6, pp. 3218–3227, June 2010.

[35] B. Dulek and S. Gezici, "Average Fisher Information maximisation in presence of cost-constrained measurements," *Electronics Letters*, vol. 47, no. 11, pp. 654–656, May 2011.

[36] R. Dwyer and L. Kurz, "Characterizing partition detectors with stationary and quasi-stationary Markov dependent data," *Information Theory, IEEE Transactions on*, vol. 32, no. 4, pp. 471–482, 1986.

[37] S. Eguchi and J. Copas, "Interpreting kullback–leibler divergence with the neyman–pearson lemma," *Journal of Multivariate Analysis*, vol. 97, no. 9, pp. 2034–2040, 2006.

[38] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation," *Econometrica: Journal of the Econometric Society*, pp. 987–1007, 1982.

[39] J. Fang and H. Li, "Power Constrained Distributed Estimation With Correlated Sensor Data," *Signal Processing, IEEE Transactions on*, vol. 57, no. 8, pp. 3292–3297, 2009.

[40] P. A. Forero, A. Cano, and G. B. Giannakis, "Distributed feature-based modulation classification using wireless sensor networks," in *Military Communications Conference, MILCOM 2008. IEEE*, 2008, pp. 1–7.

[41] F. Ghassemi and V. Krishnamurthy, "A Cooperative Game-Theoretic Measurement Allocation Algorithm for Localization in Unattended Ground Sensor Networks," *Information Fusion*, pp. 1–7, Jun. 2008.

[42] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*, 4th ed. New York, NY: Marcel Dekker, 2003.

[43] B. Guo, Q. Guan, F. Yu, S. Jiang, and V. Leung, "Energy-efficient topology control with selective diversity in cooperative wireless ad hoc networks," in *Communications (ICC), 2014 IEEE International Conference on*, June 2014, pp. 197–202.

[44] ——, "Energy-efficient topology control with selective diversity in cooperative wireless ad hoc networks: A game-theoretic approach," *Wireless Communications, IEEE Transactions on*, vol. 13, no. 11, pp. 6484–6495, Nov 2014.

[45] S. Gutta and Q. Cheng, "Data-based distributed classification and its performance analysis," in *Information Fusion (FUSION), 2012 15th International Conference on*, July 2012, pp. 1519–1526.

[46] F. Hameed, O. Dobre, and D. Popescu, "On the likelihood-based approach to modulation classification," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 12, pp. 5884–5892, December 2009.

[47] M. H. Hansen and B. Yu, "Model selection and the principle of minimum description length," *Journal of the American Statistical Association*, vol. 96, no. 454, pp. 746–774, 2001.

[48] I. Hoballah and P. Varshney, "Distributed bayesian signal detection," *Information Theory, IEEE Transactions on*, vol. 35, no. 5, pp. 995–1000, Sep 1989.

[49] K. M. Houston and D. P. McGaffigan, "Spectrum analysis techniques for personnel detection using seismic sensors," in *Unattended Ground Sensor Technologies and Applications V*, E. M. Carapezza, Ed., vol. 5090, no. 1. SPIE, 2003, pp. 162–173.

[50] S. G. Iyengar, R. Niu., and P. K. Varshney, "Fusing Dependent Decisions for Hypothesis Testing With Heterogeneous Sensors," *Signal Processing, IEEE Transactions on*, vol. 60, no. 9, pp. 4888 – 4897, Sep 2012.

[51] S. G. Iyengar, P. K. Varshney, and T. Damarla, "A Parametric Copula Based Framework for Hypotheses Testing using Heterogeneous Data," *Signal Processing, IEEE Transactions on*, vol. 59, no. 5, pp. 2308 – 2319, May 2011.

[52] S. G. Iyengar, "Decision Making with Heterogeneous Sensors - A Copula Based Approach," Ph.D. dissertation, Syracuse University, Syracuse, NY, August 2011.

[53] S. Iyengar, R. Niu, and P. Varshney, "Fusing dependent decisions for hypothesis testing with heterogeneous sensors," *Signal Processing, IEEE Transactions on*, vol. 60, no. 9, pp. 4888–4897, Sept 2012.

[54] S. Iyengar, P. Varshney, and T. Damarla, "A parametric copula-based framework for hypothesis testing using heterogeneous data," *Signal Processing, IEEE Transactions on*, vol. 59, no. 5, pp. 2308–2319, May 2011.

[55] R. Jiang and B. Chen, "Fusion of censored decisions in wireless sensor networks," *Wireless Communications, IEEE Transactions on*, vol. 4, no. 6, pp. 2668–2673, 2005.

[56] R. Jiang, Y. Lin, B. Chen, and B. Suter, "Distributed sensor censoring for detection in sensor networks under communication constraints," in *Signals, Systems and Computers, 2005. Conference Record of the Thirty-Ninth Asilomar Conference on*, October 2005, pp. 946–950.

[57] M. Kafai and B. Bhanu, "Dynamic Bayesian Networks for Vehicle Classification in Video," *Industrial Informatics, IEEE Transactions on*, vol. 8, no. 1, pp. 100–109, 2012.

[58] M. Kam, Q. Zhu, and W. S. Gray, "Optimal data fusion of correlated local decisions in multiple sensor detection systems," *Aerospace and Electronic Systems, IEEE Transactions on* , vol. 28, no. 3, pp. 916–920, 1992.

[59] V. Kapnadak and E. Coyle, "Optimal density of sensors for distributed detection in single-hop wireless sensor networks," in *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, July 2011, pp. 1–8.

[60] S. Kar, P. K. Varshney, and H. Chen, "Spatial whitening framework for distributed estimation," in *Proc. 4th IEEE Int Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) Workshop*, 2011, pp. 293–296.

[61] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, New Jersey: Prentice-Hall Inc., 1998.

[62] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, December 1971.

[63] E. Kidron, Y. Y. Schechner, and M. Elad, "Cross-Modal Localization via Sparsity," *Signal Processing, IEEE Transactions on*, vol. 55, no. 4, pp. 1390–1404, 2007.

[64] A. Krasnopeev, J.-J. Xiao, and Z.-Q. Luo, "Minimum Energy Decentralized Estimation in a Wireless Sensor Network with Correlated Sensor Noises," *EURASIP Journal on Wireless Communications and Networking*, vol. 2005, no. 4, pp. 473–482, 2005.

[65] D. Kurowicka and R. Cooke, *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley, 2006.

[66] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 17–29, 2002.

[67] J. Li and P. Stoica, "MIMO Radar with Colocated Antennas," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 106–114, 2007.

[68] S. Marano, V. Matta, and P. Willett, "Distributed detection with censoring sensors under physical layer secrecy," *Signal Processing, IEEE Transactions on*, vol. 57, no. 5, pp. 1976–1986, May 2009.

[69] D. Mari and S. Kotz, *Correlation and Dependence*. Imperial College Press, 2001.

[70] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis of nonstationary processes," *Acoustics, Speech, and Signal Processing, IEEE Transactions on*, vol. 33, no. 6, pp. 1461–1470, 1985.

[71] G. Mercier, G. Moser, and S. Serpico, "Conditional copula for change detection on heterogeneous SAR data," in *Proc. IEEE International Geoscience and Remote Sensing Symposium IGARSS 2007*, Jul. 23–28, 2007, pp. 2394–2397.

[72] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2011, vol. 526.

[73] B. T. Morris and M. M. Trivedi, "A Survey of Vision-Based Trajectory Learning and Analysis for Surveillance," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, no. 8, pp. 1114–1127, 2008.

[74] E. Msechu and G. Giannakis, "Sensor-centric data reduction for estimation with wsns via censoring and quantization," *Signal Processing, IEEE Transactions on*, vol. 60, no. 1, pp. 400–414, 2012.

[75] R. Nelsen, *An Introduction to Copulas*, 2nd ed. New York: Springer, 2006.

[76] R. Niu and P. K. Varshney, "Target location estimation in sensor networks with quantized data," *Signal Processing, IEEE Transactions on*, vol. 54, no. 12, pp. 4519–4528, 2006.

[77] R. Niu and P. Varshney, "Sampling schemes for sequential detection with dependent observations," *Signal Processing, IEEE Transactions on*, vol. 58, no. 3, pp. 1469–1481, March 2010.

[78] R. Olfati-Saber and N. F. Sandell, "Distributed tracking in sensor networks with limited sensing range," in *Proc. American Control Conf*, 2008, pp. 3157–3162.

[79] O. Ozdemir, R. Li, and P. Varshney, "Hybrid maximum likelihood modulation classification using multiple radios," *Communications Letters, IEEE*, vol. 17, no. 10, pp. 1889–1892, October 2013.

[80] O. Ozdemir, P. K. Varshney, W. Su, and A. L. Drozd, "Asymptotic properties of likelihood based linear modulation classification systems," *arXiv preprint arXiv:1211.6631*, 2012.

[81] P. Panagiotou, A. Anastasopoulos, and A. Polydoros, "Likelihood ratio tests for modulation classification," in *MILCOM 2000. 21st Century Military Communications Conference Proceedings*, vol. 2, 2000, pp. 670–674 vol.2.

[82] A. Patton, "Copula methods for forecasting multivariate time series," *Handbook of economic forecasting*, vol. 2, pp. 899–960, 2012.

[83] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: a low-communication-rate scheme for distributed detection," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 32, no. 2, pp. 554–568, 1996.

[84] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor Networks-part I: Gaussian case," *Signal Processing, IEEE Transactions on*, vol. 54, no. 3, pp. 1131–1143, 2006.

[85] A. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics*, 1st ed., ser. International Series on Biometrics. New York, NY: Springer, 2006.

[86] W. Saad, Z. Han, M. Debbah, A. Hj£rungnes, and T. Basar, "Coalitional Games for Distributed Collaborative Spectrum Sensing in Cognitive Radio Networks," *IEEE International Conference on Computer Communications (INFOCOM)*, pp. 2114–2122, Apr. 2009.

[87] W. Saad, Z. Han, M. Debbah, A. Hjorungnes, and T. Basar, "Coalitional Game Theory for Communication Networks," *Signal Processing Magazine, IEEE*, vol. 26, pp. 77–97, 2009.

[88] W. Saad, Z. Han, H. Poor, T. Basar, and J. B. Song, "A cooperative Bayesian nonparametric framework for primary user activity monitoring in cognitive radio networks," *Selected Areas in Communications, IEEE Journal on*, vol. 30, no. 9, pp. 1815–1822, October 2012.

[89] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: Real-time event detection by social sensors," in *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW '10. New York, NY, USA: ACM, 2010, pp. 851–860.

[90] T. Sandholm, K. Larson, M. Andersson, O. Shehory, and F. Tohme, "Coalition structure generation with worst case guarantees," *Artificial Intelligence*, vol. 111, no. 1, pp. 209 – 238, 1999.

[91] X. Shen, Y. Zhu, L. He, and Z. You, "A near-optimal iterative algorithm via alternately optimizing sensor and fusion rules in distributed decision systems," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 47, no. 4, pp. 2514–2529, OCTOBER 2011.

[92] J. Sills, "Maximum-likelihood modulation classification for psk/qam," in *Military Communications Conference Proceedings, 1999. MILCOM 1999. IEEE*, vol. 1, 1999, pp. 217–220 vol.1.

[93] M. Slaney and M. Covelle, "FaceSync: A Linear Operator for Measuring Synchronization of Video Facial Images and Audio Tracks," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 13, 2000, pp. 814–820.

[94] S.-H. Son, S. R. Kulkarni, S. C. Schwartz, and M. Roan, "Communication-estimation Tradeoffs in Wireless Sensor Networks," *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, vol. 5, pp. V/1065–V/1068, 2005.

[95] W. Su and J. Kosinski, "Framework of network centric signal sensing for automatic modulation classification," in *Networking, Sensing and Control (ICNSC), 2010 International Conference on*, April 2010, pp. 534–539.

[96] A. Subramanian, A. Sundaresan, and P. Varshney, "Detection of dependent heavy-tailed signals," *Signal Processing, IEEE Transactions on*, vol. 63, no. 11, pp. 2790-2803, 2015.

[97] A. Subramanian, A. Sundaresan, and P. K. Varshney, "Fusion for the detection of dependent signals using multivariate copulas," in *Proc. 14th Int Information Fusion (FUSION) Conf*, 2011, pp. 1–8.

[98] A. Subramanian, A. Sundaresan, and P. Varshney, "Fusion for the detection of dependent signals using multivariate copulas," in *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, July 2011, pp. 1–8.

[99] A. Sundaresan and P. K. Varshney, "Location Estimation of a Random Signal Source Based on Correlated Sensor Observations," *Signal Processing, IEEE Transactions on*, vol. 59, no. 2, pp. 787–799, 2011.

[100] A. Sundaresan, P. K. Varshney, and N. S. V. Rao, "Copula-Based Fusion of Correlated Decisions," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 47, no. 1, pp. 454–471, 2011.

[101] A. Sundaresan and P. Varshney, "Location estimation of a random signal source based on correlated sensor observations," *Signal Processing, IEEE Transactions on*, vol. 59, no. 2, pp. 787–799, Feb 2011.

[102] A. Sundaresan, P. Varshney, and N. S. V. Rao, "Copula-based fusion of correlated decisions," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 47, no. 1, pp. 454–471, January 2011.

[103] A. Sundaresan, "Detection and Location Estimation of a Random Signal Source Using Sensor Networks," Ph.D. dissertation, Syracuse University, Syracuse, NY, December 2010.

[104] A. Swami, Q. Zhao, Y.-W. Hong, and L. Tong, Eds., *Wireless Sensor Networks: Signal Processing and Communications*. Wiley, 2007.

[105] Z.-B. Tang, K. Pattipati, and D. Kleinman, "An algorithm for determining the decision thresholds in a distributed detection problem," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 21, no. 1, pp. 231–237, Jan 1991.

[106] W. P. Tay, J. Tsitsiklis, and M. Win, "Asymptotic performance of a censoring sensor network," *Information Theory, IEEE Transactions on*, vol. 53, no. 11, pp. 4191–4209, Nov 2007.

[107] H. L. V. Trees, *Optimum Array Processing (Detection, Estimation, and Modulation Theory, Part IV)*. Wiley-Interscience, 2002.

[108] P. K. Trivedi and D. M. Zimmer, "Copula Modeling: An Introduction for Practitioners," *Foundations and Trends in Econometrics*, vol. 1, no. 1, pp. 1–111, 2005.

[109] R. S. Tsay, *Analysis of financial time series*. John Wiley & Sons, 2005, vol. 543.

[110] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, ser. Wiley series in telecommunications. Cambridge University Press, 2005.

[111] J. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *Automatic Control, IEEE Transactions on*, vol. 30, no. 5, pp. 440–446, 1985.

[112] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing*. JAI Press, 1993, pp. 297–344.

[113] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment." *ICWSM*, vol. 10, pp. 178–185, 2010.

[114] P. Varshney, *Distributed Detection and Data Fusion*, ser. Signal Processing & Digital Filtering. Springer Verlag, 1997.

[115] M. C. Vuran, O. B. Akan, and I. F. Akyildiz, "Spatio-Temporal Correlation: Theory and Applications for Wireless Sensor Networks," *Computer Networks Journal (Elsevier)*, vol. 45, pp. 245–259, 2004.

[116] D. Wang, L. Kaplan, H. Le, and T. Abdelzaher, "On truth discovery in social sensing: A maximum likelihood estimation approach," in *Proceedings of the 11th International Conference on Information Processing in Sensor Networks*, ser. IPSN '12. New York, NY, USA: ACM, 2012, pp. 233–244.

[117] T.-Y. Wang, Y. Han, B. Chen, and P. Varshney, "A combined decision fusion and channel coding scheme for distributed fault-tolerant classification in wireless sensor networks," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 7, pp. 1695–1705, July 2006.

[118] X. Wang, M. S. Gerber, and D. E. Brown, "Automatic crime prediction using events extracted from twitter posts," in *Social Computing, Behavioral-Cultural Modeling and Prediction*, Springer, 2012, pp. 231–238.

[119] L. Wasserman, *All of Nonparametric Statistics (Springer Texts in Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[120] G. Whipps, E. Ertin, and R. Moses, "Distributed detection with collisions in a random, single-hop wireless sensor network," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 4603–4607.

[121] B. Widrow, "A study of rough amplitude quantization by means of nyquist sampling theory," *Circuit Theory, IRE Transactions on*, vol. 3, no. 4, pp. 266–276, 1956.

[122] B. Widrow, I. Kollar, and M.-C. Liu, "Statistical theory of quantization," *Instrumentation and Measurement, IEEE Transactions on*, vol. 45, no. 2, pp. 353–361, 1996.

[123] P. Willett, P. F. Swaszek, and R. S. Blum, "The good, bad and ugly: distributed detection of a known signal in dependent Gaussian noise," *Signal Processing, IEEE Transactions on*, vol. 48, no. 12, pp. 3266–3279, 2000.

[124] J. L. Xu, W. Su, and M. Zhou, "Distributed automatic modulation classification with multiple sensors," *Sensors Journal, IEEE*, vol. 10, no. 11, pp. 1779–1785, 2010.

[125] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, pp. 1–45, December 2006, article 13.

[126] H. Yoon and H. Sompolinsky, "The Effect of Correlations on the Fisher Information of Population Codes," *In*. MIT Press, 1998, pp. 167–173.

[127] C.-T. Yu and P. Varshney, "A paradigm for distributed detection under communication constraints," in *Information Theory, 1995. Proceedings., 1995 IEEE International Symposium on*, Sep 1995.

[128] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 1, pp. 39–58, 2009.

[129] Y. Zhang, N. Ansari, and W. Su, "Optimal decision fusion based automatic modulation classification by using wireless sensor networks in multipath fading channel," in *Global Telecommunications Conference, 2011 IEEE*, 2011, pp. 1–5.

[130] Y. Zheng, R. Niu, and P. Varshney, "Sequential bayesian estimation with censored data for multi-sensor systems," *Signal Processing, IEEE Transactions on*, vol. 62, no. 10, pp. 2626-2641, 2014.

[131] Y. Zhu, R. Blum, Z.-Q. Luo, and K. M. Wong, "Unexpected properties and optimum-distributed sensor detectors for dependent observation cases," *Automatic Control, IEEE Transactions on*, vol. 45, no. 1, pp. 62–72, Jan 2000.

[132] Z. Zhu and T. S. Huang, Eds., *Multimodal Surveillance: Sensors, Algorithms, and Systems*. Norwood, MA: Artech House, 2007.

VITA

NAME OF AUTHOR: Hao He

PLACE OF BIRTH: Cangzhou, Hebei, China

DATE OF BIRTH: January 29, 1989

EDUCATION: B.S. May 2010 Harbin Institute of Technology (HIT), China

PROFESSIONAL EXPERIENCE:

- Graduate Research Assistant, Sensor Fusion Laboratory, Dept. of Electrical Engineering and Computer Science, Syracuse University, Aug. 2010 - Dec. 2015

PUBLICATIONS:

1. Aditya Vempaty, Hao He, Biao Chen, and Pramod K. Varshney, "Distributed Bayesian Estimation in Sensor Networks," *Signal Processing, IEEE Transactions on*, vol. 62, no. 20, pp. 5359-5369, Oct. 15, 2014.
2. Hao He, and Pramod K. Varshney, "Fusing Censored Dependent Data for Distributed Detection," *Signal Processing, IEEE Transactions on*, vol 63, no. 16, pp. 4385-4395, Aug. 15, 2015.
3. Hao He, and Pramod K. Varshney, "A Coalitional Game for Distributed Inference in Sensor Networks with Dependent Observations," to appear in *Signal Processing, IEEE Transactions on.*

4. Hao He, Arun Subramanian, Pramod K. Varshney, and Thyagaraju Damarla, “Fusing Heterogeneous Data for Detection Under Non-stationary Dependence," in *Proc. 15th International Conference on Information Fusion*, 2012.

5. Hao He, Arun Subramanian, Xiaojing Shen, and Pramod K. Varshney, “A Coalitional Game for Distributed Estimation in Wireless Sensor Networks," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013.

6. Hao He, and Pramod K. Varshney, “Distributed Detection with Censoring Sensors under Dependent Observations," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014.

7. Hao He, Sora Choi, Pramod K. Varshney, and Wei Su, “Distributed Classification under Statistical Dependence with Application to Automatic Modulation Classification," in *Proc. 18th International Conference on Information Fusion*, 2015.

8. Hao He, Arun Subramanian, Sora Choi, Pramod K. Varshney, and Thyagaraju Damarla, “Social Media Data Assisted Inference with Application to Stock Prediction," to appear in *Proc. IEEE Asilomar Conference on Signals Systems and Computer*, 2015.

9. Sora Choi, Hao He, and Pramod K. Varshney, "Copula-based Dependence Modeling for Inference in Radar Systems," to appear in *Proc. IEEE Radar Conference*, 2015.

10. Satish Iyengar, Hao He, Arun Subramanian, Ruixin Niu, and Pramod K. Varshney, "Distributed Detection and Data Fusion with Heterogeneous Sensors," *Multisensor Data Fusion: From Algorithm and Architecture Design to Applications*, CRC Press, 2015.